DOCUMENT RESUME

ED 038 153 52 LI 001 903

AUTHOR	Cady, Glee; And Others
TITLE	System Scope for Library Automation and Generalized Information Storage and Retrieval at Stanford University.
INSTITUTION	Stanford Univ., Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-7-1145
PUB DATE	Feb 70
GRANT	OEG-1-7-071145-4428(095)
NOTE	152p.
EDRS PRICE	EDRS Price MF-$0.75 HC-$7.70
DESCRIPTORS	*Automation, Decentralized Library Systems, *Electronic Data Processing, *Information Retrieval, *Information Science, *Information Storage, Information Systems, Time Sharing, University Libraries
IDENTIFIERS	BALLOTS, *Bibliographic Automation Large Library Operations, SPIRES, Stanford Physics Information Retrieval System

ABSTRACT

The scope of a manual-automated system serving the 40 libraries and the teaching and research community of Stanford University is defined. Also defined are the library operations to be supported and the bibliographic information storage and retrieval capabilities to be provided in the system. Two major projects have been working jointly on library automation and information retrieval since 1968. One is the Bibliographic Automation of Large Library Operations on a Time-sharing System (BALLOTS) funded by the Office of Education and the other is the Stanford Physics Information Retrieval System (SPIRES), funded by the National Science Foundation. The creation of a production system for library automation (BALLOTS II) and generalized information storage and retrieval (SPIRES II) requires the continuation of a comprehensive system development process. This process has six phases: (1) preliminary analysis, (2) detailed analysis, (3) general design, (4) detailed design, (5) implementation and (6) installation. The document represents the main output of the preliminary analysis phase encompassing the definition of goals, description of the user environment, analysis of the existing system, selection of the system scope and establishment of gross technical feasibility of the selected first implementation scope. Included is a 20-page glossary of information science terminology. (MF)

ED038153

OE/BR-7-1145
PA-52

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

LI 001903

SYSTEM SCOPE
for
LIBRARY AUTOMATION
and
GENERALIZED INFORMATION STORAGE AND RETRIEVAL
at
STANFORD UNIVERSITY

LI 001903

LI 001903

February, 1970
Stanford University
Stanford, California 94305

ACKNOWLEDGMENT

The research reported herein was performed pursuant to:

Grant No. OEG-1-7-071145-4428 (095) with the Office of Education, United States Department of Health, Education and Welfare (BALLOTS -- Bibliographic Automation of Large Library Operations Using a Time-Sharing System), and

Grant Nos. GN 600, GN 742 and GN 830 with the National Science Foundation (SPIRES -- Stanford Physics Information Retrieval System).

Grantees undertaking such projects are encouraged to express freely their professional judgement in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official agency position or policy.

CONTRIBUTORS

This Scope Document was written by the SPIRES/BALLOTS project staff: Glee Cady, Jesse Caton, Clark Crane, Wayne Davison, Hank Epstein, Douglas Ferguson, Carol Kayser, Cheryl Kerr, James Marsheck, Donn Martin, Thomas Martin, Eleanor Montague, William Riddle, John Schroeder, and Jerrold West.

It is the result of work by many individuals. We cannot mention all who have contributed but it is a pleasure for us to express our special appreciation to the following:

Professor William F. Miller, Vice President for Research and Chairman of the SPIRES/BALLOTS Executive Committee has been a constant source of encouragement. We have also been aided in many ways by his Special Assistant, Nicholas Fortis.

Much of our work has required the support of the Stanford Computation Center. Paul Armer, Director of the Stanford Computation Center, has been extremely helpful with ideas and suggestions. Rod Frederickson, Associate Director of the Campus Facility, has given extensive technical support.

We particularly acknowledge the critical comments and ideas of Mr. Allen B. Veaner, Assistant Director for Bibliographic Operations, Stanford University Libraries, Principal Investigator for the Office of Education BALLOTS project, and Professor Edwin B. Parker, Institute for Communication Research, Principal Investigator for the National Science Foundation SPIRES project.

The staff of the various libraries of Stanford University made major contributions to this Document. David C. Weber, Director of Libraries, has given concrete and continuous support for the contributions which library automation can make to improved library management. The Analysis of the current library system and the scope definition could not have been completed without the active support of Robert Golter, Assistant Director for Undergraduate and Branch Services, and of the following department heads: Judy Fair, Government Documents; Ralph Hansen, Acquisition; Jennette Hitchcock, Catalog; Jack Plotkin, Central Circulation; and B. Jack Pooler, Science Libraries.

The staff members of each of these departments contributed their time and detailed knowledge of operations to the Systems Analysis. We particularly acknowledge the efforts of William Allan, Guy DeBall, Kay Cutler, Pamela Dempsey, Ann Dietz, Charles Gorham, Peter Johnson, Fred Lynden, Maria Nikonenko, Frances Peters, Richard Pollard, and Gertrude White. Bernard J. Denham, Assistant to to the Director of Libraries, made extensive statistical data available to us.

At the Stanford Linear Accelerator (SLAC) Library, George Owens and Robert Gex gave whole-hearted support to the initiation fo the prototype system. Louise Addis did much of the background work in setting up the system and is primarily responsible for its effective use. In addition, she wrote Appendix F which records the SLAC experience with the prototype automated system.

A Preliminary Version of the Scope Document was comprehensively criticized by many individuals previously mentioned. In addition, we acknowledge pertinent comments from the following: Professor J. Myron Jacobstein, Librarian, Stanford Law Library; Karen Bendorf, Catalog Department; and John Heyeck, Assistant to the Director of Libraries.

Diana DeLanoy and Hilma Mortell, who are no longer with the project, contributed to several sections of a preliminary version of the Scope Document.

The flowcharts in the text were skillfully drawn by Sandra Beattie and many of the details of reproduction and distribution were handled expertly by Naomi Schwartz. Finally, we wish to thank the secretarial staff who contributed in many ways, each day, as the document was being prepared: Marlene Amiot, Libby Jenkins and Charla Meyer. +

TABLE OF CONTENTS

i Acknowledgement

ii Contributors

CHAPTER PAGE

PART I
Preface To the System Scope

1.0 INTRODUCTION........ 2
- 1.1 Project Rationale........ 2
- 1.2 Purpose and Audience 3
- 1.3 Document Organization 3
- 1.4 Suggestions to the Reader........ 4

2.0 BACKGROUND........ 4
- 2.1 BALLOTS I and SPIRES I 4
- 2.2 A Perspective on Development--BALLOTS II and SPIRES II........ 6

3.0 GOALS 8
- 3.1 Library Automation--BALLOTS........ 8
- 3.2 Generalized Information Storage and Retrieval--SPIRES........ 9
- 3.3 Common Facilities--BALLOTS and SPIRES... 10

PART II
Library Automation

4.0 CURRENT LIBRARY SYSTEMS AND THEIR LIMITATIONS
- 4.1 Users and User Characteristics........ 12
- 4.2 Summary of Library Operations........ 13
  - 4.21 Technical Processing 14
  - 4.22 User Services........ 19
  - 4.23 Summary of Limitations........ 23

5.0 LONG RANGE SCOPE FOR LIBRARY AUTOMATION 25
- 5.1 General Considerations 25
  - 5.11 Technical Processing 25
  - 5.12 User Services........ 25

6.0 FIRST IMPLEMENTATION SCOPE........ 28
- 6.1 Technical Processing........ 28
- 6.2 User Services........ 35

PART III
Generalized Information Storage and Retrieval

7.0 CURRENT STATUS, GENERALIZED INFORMATION STORAGE AND RETRIEVAL 38
- 7.1 Representative User Profiles 38
- 7.2 Summary of User Requirements 40
- 7.3 Summary of Current Facilities and Limitations........................ 41

8.0 LONG RANGE SCOPE FOR GENERALIZED INFORMATION STORAGE AND RETRIEVAL 45
- 8.1 Retrieval 45
- 8.2 File Management 54

9.0 FIRST IMPLEMENTATION SCOPE.................... 60
- 9.1 Retrieval 60
- 9.2 File Management 62

PART IV
Shared Facilities

10.0 SUMMARY OF CURRENT SHARED FACILITIES.......... 66

11.0 LONG RANGE SCOPE, SHARED FACILITIES........... 69

12.0 FIRST IMPLEMENTATION SCOPE, SHARED FACILITIES. 69

APPENDICES

APPENDIX A Glossary 74
APPENDIX B Preliminary Analysis Phase Methodology..... 95
APPENDIX C Documentation of the Current Library System 99
APPENDIX D The Libraries of Stanford University 113
APPENDIX E The Stanford Law Library--A Potential BALLOTS and SPIRES User.................... 116
APPENDIX F Stanford Linear Accelerator Center Participation in SPIRES.................... 122
APPENDIX G Tutorial: Information Storage and Retrieval 141

LIST OF FIGURES

Figure 1: Acquisition First Implementation Scope...... 30

Figure 2: Cataloging First Implementation Scope 33

PART I

PREFACE TO THE SYSTEM SCOPE

1.0 INTRODUCTION TO THE SCOPE DOCUMENT

1.1 Project Rationale

Library automation requires a major system development effort and sizeable expenditures for computer equipment. Computerized information storage and retrieval requires an equally large investment in hardware and software. Both efforts have common conceptual problems in such areas as bibliographic file organization and on-line searching. Each effort derives benefits from the other. Bibliographic files created in the process of library automation are available for generalized retrieval uses, and complex retrieval routines are available for search of library bibliographic files.

At Stanford, two major projects have been working jointly on library automation and information retrieval since 1968. One is BALLOTS (Bibliographic Automation of Large Library Operations on a Time-sharing System), funded by the Office of Education and the other is SPIRES (Stanford Physics Information REtrieval System--informally known as the Stanford Public Information REtrieval System), funded by the National Science Foundation. The purpose of this collaboration is to create the common software required to support both the BALLOTS and the SPIRES applications. The joint effort is overseen by the SPIRES/BALLOTS Executive Commitee chaired by Professor William F Miller, Vice-President for Research. Other members of the Executive Commitee are: David C. Weber, Director of Libraries, Paul Armer Director of the Stanford Computation Center, Professor Edwin B. Parker, Institute for Communication Research and Principal Investigator for SPIRES, and Allen B. Veaner, Assistant Director of University Libraries for Bibliographic Operations and Principal Investigator for BALLOTS.

The Stanford project structure and system development philosophy reflect the common uses and individual needs of both BALLOTS and SPIRES. The concept of shared facilities, described in detail in Part IV, refers to the systems software and hardware designed to service both the BALLOTS application and the SPIRES application. Examples are, an on-line text editor and a computer terminal handler. Both are shared software facilities which can service bibliographic input and specialized research files. Computer hardware such as a central processing unit or direct access devices (allowing shared files) are examples of shared hardware facilities. Combining resources in this system development effort reduces the cost of creating common facilities and provides a pool of skilled manpower resources for each area. The approach of specialized applications based on shared facilities is reflected in the organization of this scope document.

1.2 Purpose and Audience

The purpose of this document is to define the scope of a manual-automated system to serve the libraries and the teaching and research community of Stanford University. The scope sets the limits and focuses the activity of the system development effort. The managers of a large research library are aware of the pressing needs of their organization and its patrons. Those who are responsible for planning and allocating computer resources to meet the educational and research needs of the university know the complexity of this task. This document defines the library operations to be supported, and the bibliographic information storage and retrieval capabilities to be provided in the system. It is directed to librarians who will use the system, to research and computer personnel who are developing it, and to university administrators and directors of libraries who need to make the policy decisions on the installation of such a system.

1.3 Document Organization

The Scope Document is organized into four parts followed by appendices which provide supplemental and supporting information. Part I introduces the document, gives the project background and states the goals of the library automation (BALLOTS) and information storage and retrieval effort (SPIRES) at Stanford.

Part II summarizes the current library system, its limitations and the scope of computer services which will deal with these limitations. A subset of the total required services is selected (BALLOTS II) and presented for implementation during the current system development cycle.

Part III discusses the users, requirements and status of the current generalized information storage and retrieval system (SPIRES I) at Stanford. Limitations are described and a long range scope of activities is proposed to deal with these limitations. A selection is made from this scope of activities (SPIRES II) for implementation in the current development cycle.

Part IV discusses Shared Facilities. Essentially these are the hardware and software required to service both specialized library applications (BALLOTS) and the generalized information storage and retrieval applications (SPIRES).

1.4 Suggestions to the Reader

A discussion of library operations and computer functions inevitably involves the presentation of material at varying levels of technical complexity. Specialized terminology familiar either to librarians or computer professionals is often not familiar to the other. Every attempt has been made to communicate with a minimum of technical terminology. Those who are not conversant with the concepts of computerized information storage and retrieval will find Appendix G (Tutorial: Information Storage and Retrieval) helpful. Technical terms in both the library and computer fields are defined in Appendix A (Glossary).

Readers whose interest is oriented toward libraries will find Part II of greatest interest and readers oriented toward computer information systems will find Part III of greatest interest. Both groups are advised to read Part I Chapter 3.0 GOALS, carefully since it gives the overall direction of development, and the discussion of Shared Facilities (Part IV) since these serve all bibliographic and retrieval applications.

2.0 BACKGROUND

2.1 BALLOTS I and SPIRES I

The publication explosion, the compelling need for access to information, and rapid library growth are not unique to Stanford University. At Stanford, a commitment has been made to deal with the information problems of the university by improving library service and developing a campus based bibliographic retrieval system. Using the tools of computing technology and library systems analysis, computer specialists have joined with librarians and behavioral scientists in exploring the problems and creating the systems to meet the bibliographic requirements of a major university community.

Since 1967 the Stanford University Libraries and the Institute for Communication Research have conducted research projects with funding from the Office of Education (BALLOTS) and the National Science Foundation (SPIRES) respectively. In 1968 the shared perspective and close collaboration of these two projects was formalized by placing them under the SPIRES/BALLOTS Executive Committee chaired by Professor William F. Miller, Associate Provost for Computing and Vice-President for Research.

Stanford University was an appropriate setting to initiate research and development in bibliographic retrieval.

Interest in automation was strong in all areas of the Stanford University Libraries and especially with its Director (then Associate Director), David C. Weber, and Assistant Director for Bibliographic Operations, Allen B. Veaner. The library had achieved during 1964-66 a remarkably successful computer produced book catalog for the J. Henry Meyer (Undergraduate) Library. Professor Edwin B. Parker and his colleagues at the Institute for Communication Research were already applying to computer systems the behavioral science analysis which had previously been applied to print, film and television media. The Stanford Computation Center, under Paul Armer, had at its Campus Facility a powerful IBM 360 model 67 computer, a locally developed time sharing system and a first rate programming staff associated with one of the nation's leading Computer Science departments. A close working relationship between the University Libraries, the Computation Center, and the Institute for Communication Research was the firm foundation for research and development.

The project software development group applied itself to writing programs necessary for bibliographic retrieval. In the Library, an analysis and design group worked closely with the library staff in studying library processes and defining requirements. This joint effort created a prototype system which could be used in the main library and by Stanford faculty and students, primarily high energy physicists.

In early 1969, two prototype applications were activated using the jointly developed systems software; an acquisition system was established in the Main Library (BALLOTS I) and a bibliographic retrieval system (SPIRES I) was established for a group of High Energy Physicists. Centralized management of library input was handled by two newly created departments, Data Preparation and Data Control. In the library, several terminals were installed for on-line searching. An on-line In Process File was created consisting of 30% of the Roman alphabet acquisition material ordered by the library. On-line searching was conducted daily during regular library hours by a specially trained staff. This prototype system operated during most of 1969, demonstrating the technical feasibility of the combined project goals. It was studied and evaluated by the library systems and programming staffs. A great deal was learned about the human, economic and technical requirements of a bibliographic retrieval system. Part II and Part III of this document summarize some of this evaluation and show the relation of these findings to the scope of a production system.

2.2 A Perspective on Development--BALLOTS II AND SPIRES II

The result of operating the prototype applications (BALLOTS I and SPIRES I) was very encouraging, particularly with respect to the advantages of utilizing common software. Feasibility and usefulness were clearly established and a wealth of knowledge was gained under actual operating conditions. The joining of library and retrieval application areas served by Shared Facilities (hardware and software) was shown to be a rewarding approach.

BALLOTS I and SPIRES I resulted from a development process in which user requirements were analyzed, programs written and tested, and prototypes created and evaluated. Librarians, behavioral scientists, library systems specialists and computer specialists collaborated over an extended period of time. The development process which produced the successful prototype system was a major milestone. The outcome was the definition of a production bibliographic retrieval system with distinctive hardware and software requirements.

The creation of a production system for library automation (BALLOTS II) and generalized information storage and retrieval (SPIRES II) requires the continuation of a comprehensive System Development Process. This process is a framework within which tasks are defined, assigned and coordinated. The System Development Process for the creation of BALLOTS II and SPIRES II has six phases:

Phase A: Preliminary Analysis

Phase B: Detailed Analysis

Phase C: General Design

Phase D: Detailed Design

Phase E: Implementation

Phase F: Installation

Preliminary Analysis involves the definition of goals, description of the user environment, analysis of the existing system, selection of the system scope and establishment of gross technical feasibility of the selected first implementation scope. The Scope Document (which you are now reading) represents the main output of the Preliminary Analysis Phase.

Detailed Analysis enumerates minutely the requirements to be met by the manual-automated system. (1) Performance requirements are stated quantitatively, including response time, hours of on-line accessibility, allowable mean failure time, maximum allowable recovery time and similar factors. (2) Record input/output is determined in terms of volume, growth, and fluctuations. Timing considerations for batch input/output are determined in order to plan for scheduling requirements. (3) All input/output document formats are determined on a character by character basis. (4) Rules transforming input data elements into output data elements are formulated and tabulated, and (5) the upper bounds of development and operating costs are established.

General Design encompasses both system externals (procedures, training, reorganization, etc.) and system internals (alternative hardware and software solutions to the stated requirements). An overall software approach hardware configuration is selected and expressed in a General Design Document.

Detailed Design completes the internal and external design, creates implementation and testing plans, and provides programming specifications. These are incorporated in a Detailed Design Document.

In the Implementation Phase, user documentation is created and personnel training begins. Programs are coded and checked out, systems and pilot testing is carried out and critiqued. A variety of documents result: programs, maintenance documentation, and test results.

In the Installation Phase, training of all personnel is completed, files are converted and, after a time of parallel operation with the manual system, a changeover is made to the automated system. Performance statistics are collected and a support plan and project history are written.

Each phase description has been necessarily abbreviated. Not all activities or outputs have been described. Some of the phase activities overlap and feed back to redefine previous activities. A "Wishbook" which has been maintained through all phases is put in final form in the Installation Phase. The "Wishbook" is very important since it represents the link to successive development iterations. It contains information on capabilities, services and operational characteristics the desirability of which became apparent during the development process but which could not be included because of time, cost or technical constraints. The Wishbook also contains information on internal (programming or hardware) and external (user or procedural) operational deficiencies determined after the system has been running for some period of time. This information will be considered in designing new portions and will aid in the overall improvement of the system.

This statement of the System Development Process guides SPIRES/BALLOTS II development from the definition of goals to the installation of a fully operational system.

3.0 GOALS

This chapter presents the general objectives of the system. Goals provide a direction for activity and a standard against which to measure achievements. Specifying goals, expressing them as a series of related tasks and assessing their outcomes is a continuing activity in the system development process.

The project goals are presented as they relate to Library Automation (BALLOTS), Generalized Information Storage and Retrieval (SPIRES), and Shared Facilities. These goals are interrelated. The goals of Shared Facilities (hardware and software) support and serve the goals of BALLOTS and SPIRES.

3.1 Goals--Library Automation--BALLOTS

As the major information center of a large academic institution, the library must respond effectively and economically to the university community. The library is a complex combination of people and machines providing the major bibliographic resources of the university to students and faculty. It reflects the needs and priorities of a changing university environment. The university library is also part of a larger network of information sources which includes other research libraries, The Library of Congress and specialized information storage agencies.

The essential goals of BALLOTS are expressed in a library system (both the manual and automated portions) which is:

USER RESPONSIVE. It adapts to the changing bibliographic requirements of diverse user groups within the university community.

COST COMPETITIVE. It provides fast, efficient internal processing of increasing volumes of material. This is accomplished at unit costs which are lower than manual costs for comparable volumes of processing transactions.

SYSTEM OPTIMIZED. It is not an attempt to automate portions of the existing manual system. It is based on the actual operating requirements of library processing and is not dependent on the existing procedural, organizational or physical setting.

PERFORMANCE ORIENTED. It provides the library and university administration with data which are useful for the measurement of internal processing performance and user satisfaction.

FLEXIBILITY. It has the capability for expansion to embrace a broader range of services and a wider group of users. It will be able to link up and serve other information systems and effectively use national data sources.

These goals will be expressed in specific capabilities which will (among other things): minimize manual filing, eliminate many clerical tasks now performed by professionals and provide user suggestion mechanisms. The effect of these computer capabilities will be: to drastically reduce errors associated with manual sorting, typing and hand transcription; to speed the flow of material through library processing; to aid book selection by providing fast access to central machine files; and to enable librarians to advise a patron of the exact status of a work about which he inquiries. In summary, responsiveness to library users, efficiency of operation, optimization, performance monitoring and flexibility for future improvement, are the essential goals of library automation.

3.2 Goals--Generalized Retrieval--SPIRES

The SPIRES generalized information storage and retrieval system will support the research and teaching activities of the library, faculty, students, and staff. Each user will have the capability of defining his requirements in a way which automatically tailors the system response to his individual needs. The creation of such a system is a major activity involving the study of users, source data, record structure, file organization and considerable experimentation with facilities. The SPIRES system will be characterized by flexibility, generality and ease of use. The goals of SPIRES in specific areas are as follows:

DATA SOURCE AND CONTENT. A generalized information storage and retrieval capability will store bibliographic, scientific, administrative and other types of records in machine readable form. Collections will range from large public files converted from centrally produced machine readable data such as MARC (see Glossary) to medium-small files created from user generated input (faculty, student files).

SEARCH FACILITIES. It will provide the capability for searching files: interactively (on line) via a computer terminal, on a batch basis by grouping requests and submitting on a regular schedule or on a standing request basis in which a search query is routinely passed against certain files at specified intervals.

FEEDBACK. Reports on the use frequency of various system elements will be provided. This will include statistical analyses of user difficulties and system errors.

RECORD MODIFICATION. Update and edit capability will be provided on a batch basis or on-line; and options for update will be at the level of record, data element and character string within data element.

COSTS AND CUSTOMERS. The cost of these services should be sufficiently low for a wide range of customers to cost justify their use of the system. The variety of services should be sufficiently great to encourage a growing body of users. Costs and services must be related at various levels to permit users to select the type of service which meets their needs within the limits of their economic resources.

3.3 Goals--Shared Facilities--BALLOTS and SPIRES

Shared facilities are software and hardware designed to provide concurrent service to BALLOTS and SPIRES applications. Since the sharing of such resources represents a substantial savings to all applications served, maximum attention will be given to the sharing concept. Whenever possible, advantage will be taken of economies gained by providing major facilities for multiple applications.

HARDWARE. The hardware environment will provide reliable, economical, and flexible support to those applications residing within it.

SOFTWARE. The software, which will consist of an operating system, an on-line executive program, a terminal handler, a text editor, and many other facilities, will be jointly used by various applications.

GENERALITY/EXPANDABILITY. The shared facilities will be designed to allow growth of the current applications as well as to allow the addition of new applications to Shared Facilities without modification to previous applications.

PART II

LIBRARY AUTOMATION

4.0 CURRENT LIBRARY SYSTEMS AND THEIR LIMITATIONS

4.1 Users and User Characteristics

The Stanford Library Community is made up of over 40 libraries with a staff of 465. It serves the students, faculty, and staff of Stanford University and aids the research efforts of various industries, visiting scholars, and other educational institutions.

4.1.1 Institutional Users

The libraries of the Stanford campus consist of two groups. Most are a part of the Stanford University Libraries, headed by David C. Weber who reports directly to the Provost of the University. In addition, there are six Coordinate Libraries, each of which is headed by a librarian who reports to the Dean of the school or Director of the supporting institution. (See Appendix D for a complete list of libraries at Stanford.) These two groups are linked through the University Library Council chaired by Mr. Weber. As potential users of an automated bibliographic system these library organizations are interested primarily in increased economy and efficiency for their operations and in better service to their users. One specific area of interest is the ability to hold down unit costs in the face of increasing work loads.

4.1.2 Library Staff Users

As a user population, the library staff consists of four groups, based on their training and experience.

1. Senior librarians are a group of highly qualified, experienced, employees who make policy for and oversee major portions of library operations. Some are responsible for the selection, processing, and/or maintenance of special collections. Others are in charge of major administrative units in the library. These librarians need a system which helps them with their administrative tasks, especially by providing statistics and promoting economical operation through the control provided by better management reporting. Budgetary considerations are of prime importance to these librarians. They have many ideas and opinions which can contribute to the design of an automated library system.

2. There are other librarians whose work involves specific professional responsibilities. These librarians want a system which will free them from clerical tasks and allow them to devote more time to professional tasks.

3. Senior library assistants generally have extensive experience in a specific area of library operations, or a specific subject area, and have major supervisory and training responsibilities for supporting staff. Therefore they require a system which is extremely dependable, easy to use, and easy to teach others to use.

4. The library assistants include, among others, part-time student employees and wives of graduate students, who are typically of above average intelligence and quick to learn. It is expected that feedback from this group will contribute significantly to the continued improvement of the automated system. They may adapt more readily to a new system, especially because of their lack of preconceived ideas about how it should work.

Any employee who works with an automated system on a day-to-day basis is likely to be impatient with a complicated system which is difficult to use and takes a long time to master. The employee who is responsible for the day-to-day details of operation will especially welcome a system which speeds routine work. All who use the system will be concerned about its accuracy and reliability.

4.1.3 Patron Users

As with the library staff, the needs and experience of library patrons cover a wide range. A faculty member, researcher, or graduate student makes rigorous demands on a library for detailed and comprehensive information.

At the other extreme, an undergraduate may need basic guidance before he can even define and articulate his information requirements. The recreational user often wishes to browse through material. All these users have two things in common: (1) they have little knowledge or interest in the Library's internal processes, and (2) they expect and demand rapid, efficient service from the library. They may not use the library every day, so any aspect of a system which they use must be immediately understandable. The system must be operating and available during library service hours. And finally the system must be responsive to the individual user and provide him with messages at his level of understanding which help him to use the system in his particular situation.

4.2 Summary of Library Operations

In addition to general administration, library operations are divided into two general categories: Technical Processing and User Services. Technical Processing activities relate to building, organizing and maintaining the library's collections. These activities include the acquisition, cataloging, binding and finishing, and repair of library material. User Services relate to activities which make relevant library material available to users. They include selection of library material, circulation system operation, inter-library loan, and reference service. The following paragraphs discuss the nature and purpose of individual library operations and their present functioning in the Stanford University Libraries.

4.21 Technical Processing

A. Acquisition

Objectives, Products, and Services

The primary objective of an acquisition system is to control the flow of material from various sources into the library. The purpose of this control is to 1) maintain a record of the status of material from the point of request or receipt, through cataloging and end processing, to the stacks and 2) coordinate acquisition with user requests, available book funds, vendor arrangements and the library's holdings.

An acquisition system must accommodate a variety of different acquisition modes (for example, gift, exchange, purchase order, standing order, on approval), and various material types (for example, films, books, serials, microtexts). The system must handle wide fluctuations in work loads, several languages, and varying fund restrictions. Communication with requesters, vendors and other library departments and the maintenance of management statistics are additional requirements.

Current Methods

There is a centralized Acquisition Department which services all units of the Stanford University Libraries. Organizationally, acquistion is unified but functionally two Divisions of that Department specialize in the acquisition of serials and material by gift or exchange. A third unit, the Order Division, specializes in monograph and non-book material acquisition and also serves as a general purchasing unit for the Stanford University Libraries. In addition, there are other library divisions and departments which acquire material for their own areas (for example, the Government Documents Department). In general, the Order Division prepares all purchase orders; each acquisition unit receives its own material and invoices and approves invoices for payment. Voucher preparation and communication with the University's Accounts Payable unit is centrally handled by the library's Financial Office.

Material selection and the decision to buy, although not strictly a part of acquisition processing, are closely related. Responsibility for the selection of material rests with the librarians, faculty, Resources Development Program curators and others who use the library's acquisition system.

In general, each department or division engaged in selection or acquisition operates independently, using varying procedures. There are, however, two common features of an acquisition system: (1) The use of manual files of status records and (2) searching of manual files as a basis for decisions and action.

1. Two basic manual files are used to control material in process: an In Process or Order File arranged alphabetically by main entry and a Dealer File sequenced by Order Number. These files contain acquisition, searching, requester, fund, status and bibliographic data. To fulfill special requirements, some units divide these files into subfiles such as "filled order" and "outstanding orders" or, "filled", "unfilled" and "standing order" vendor files. Also, special purpose files are maintained to control activities such as claiming and cancelling, out of print, gift and on approval acquisition, and exchange correspondence. Files are updated as needed.

2. The basic acquisition input is data relating to a book, invoice, or purchase request. The first step is to search for the item in manual files and printed reference tools to answer questions such as:

Does the item exist?
Is the item already on order?
Is the item already in the collection?
Is an added copy wanted?
Is Library of Congress bibliographic information
  available?
Is the item out of print?
Has the material listed on an invoice been
  received?
  action required.

The results of the search dictate the action required. Searching involves human decisions, intuition and experience; its path varies with the kind of information available and the type of item being searched. The output is a document. For example:

Purchase Order, Claim, Cancellation to Vendor,
Notification of Material Receipt to Requester,
Approved invoice to Financial Office.
Acquisition and searching data to the Library
of Congress' National Program for Acquisitions
and Cataloging.

Limitations

If the acquisiton system is to be responsive to future requirements, increasing processing loads and rising costs, the limitations of the system must be identified. For example:

1. MULTIPLICATION OF MANUAL FILES

Multiple acquisition sources and the constraints of single access files result in a proliferation of files. This means increased time for searching and maintenance and in some cases, the use of inefficient manual procedures. Examples of the many files used in the Order Division are:

Order File
Meyer Order File
Dealer File (divided into three parts)
Material Received--No Invoice File
Invoice--No Material File
Overseas Order File (divided into two parts)
On Approval File

2. FILE DEGRADATION

Manual files are increasing in size, in difficulty of use and in residual error due to:

1. Insufficient purging
2. Frequent misfiling compounded by frequent refiling
3. Multiple, uncontrolled sources of hand written updates
4. Records on flimsy papers attached by staples and clips
5. Insufficient coordination of the form of an acquisition entry with corrected catalog entry, a problem which results in unnecessary duplicate ordering and searching.

3. LACK OF CONTROL INFORMATION

Current manual acquisition procedures and files cannot efficiently or economically support systematic monitoring procedures such as claiming for material or invoices. There is no economical method for monitoring the performance of over 2,000 vendors or of easily collecting and summarizing statistics on personnel productivity or departmental performance. The time consuming nature of hand written updates to manual file records makes the adequate control over items in various stages of processing.

4. RISING COSTS

Since 1964/65, Acquisition staff has increased 1.7 times, Departmental expenditures have increased 2.5 times and material processing costs have tripled.

B. Cataloging

Objectives, Products, and Services

The primary objectives of a cataloging system are to (1) describe and classify material entering the collection, (2) designate index entries (author, title, series, subject, special concepts), (3) create records for public and staff files, and (4) maintain those records and files to reflect changes, additions and deletions.

To meet these objectives, a cataloging system must be able to process material in many languages, in various subject fields, and with different formats and character sets as well as publication frequencies (for example, serials, phonorecords, microtexts). The principal output is a set of records for public and staff files, usable for finding a known item or a group of items with a common characteristic. The system must (1) make optimum use of existing bibliographic data, (2) control items while in cataloging process, (3) collect management statistics.

Current Methods

A central Catalog Department services the Stanford University Libraries. The Department is organized into functional units (1) for cataloging various categories of material such as music, Slavic, special collections, monographs, serials, overseas campuses and the Meyer Library (for which a computer produced book catalog is regularly published), (2) for records maintenance, and (3) for card preparation and production.

Within the Catalog Department, there are processing variations among the functional units. Nevertheless, there are common processing routines used throughout the department to meet the requirements of a cataloging system.

1. The basic cataloging input is an item to be added to the collection. New material is described bibliographically, classified and made accessible according to standard tools (such as the Library of Congress Classification Schedules, Library of Congress Subject Heading List and the Anglo American Cataloging Rules) and existing Library of Congress and Stanford conventions. A search procedure takes place using the Stanford shelf list and Main (Union) Catalog, Library of Congress catalogs, National Union Catalog and the Library of Congress depository card file. If the search procedure does not uncover any pre-existing bibliographic data, a record is created (original cataloging). If information is found, it is checked, and if necessary, modified to conform to Stanford conventions. The bibliographic record includes: main entry, body of the entry, collation (pagination, illustrations, size), notes,

call number, location, tracings for added entries and cross references.

2. The bibliographic record created during cataloging is duplicated making a card set and added entries, shelving and filing locations are entered on the card. Each card set is revised, sorted and distributed for filing. The Department produces cards for the main research library and departmental libraries. Cards are produced at cost for Coordinate Libraries and other agencies.

3. The library collection is dynamic and thus records and files are constantly being modified. Added copies, added volumes, transfers, discards, and changes in bibliographic and call number data must be reflected in existing records and files.

Limitations

Since 1964/65, Catalog staff has increased 2.4 times while Department expenditures have more than tripled. Unit costs for Cataloging alone have increased during this same period from about $6.70 per book to about $8.70 per book (This unit cost does not include any processing cost attributable to the Acquisition Department). Volume of titles cataloged has increased 2.4 times. Much of this serious increase in costs can be attributed to the following limitations:

1. PROLIFERATION OF MANUAL FILES

The physical separation of cataloging staff from the Main (Union) Catalog and Circulation Shelf List has necessitated the creation of separate authority files, decision files, and instruction files in the Catalog Department. The maintenance and updating of these manual files consumes personnel time; the penalty for incomplete or poor maintenance is the perpetuation of errors and increased maintenance work.

2. DILUTION OF PROFESSIONAL TASKS

The substantial distance between the Main Catalog, Serials Record, Order File, Loan Desk charge records and Government Document files necessitates a considerable amount of "walking time" for the establishment of headings, investigation of changes in records and location of material in process or in circulation.

The increased output of cataloging makes a heavier work load in the production and maintenance of catalog records which requires an increased number of assistants and an increase in supervision.

The changes in the Library of Congress classification schedules and subject heading lists poses a need for perpetual updating of these tools though this work may be unproductive of actual output.

The supervision of burgeoning routine and repetitive tasks, and the need for changes in procedures due to new developments, greatly reduces the percentage of time spent on actual cataloging by professionals. Inefficiency in space arrangements means more staff; manual files increasing in size means more staff. More staff means more supervision.

3. LACK OF CONTROL INFORMATION

Current procedures do not permit adequate control over items in process, making it difficult to (a) locate a specific book, (b) assess the current cataloging work load of an individual or unit and assess the significance of the retrospective cataloging work load (arrearages) for determination of priorities on demand. Management statistics are often insufficient because they do not give the variety of processing breakdowns and costs which are required to evaluate Department performance.

4.22 User Services

4.22.1 CIRCULATION

Objectives, products, and services

The principal objective of a circulation system is to make library materials available in an equitable and efficient manner. To accomodate the needs of the user, the nature of different materials, and current demands for the material, the library sets differing loan periods for differing types of borrowers and differing types of material. The Library must also maintain information about all items in circulation, including identification of the material, name of the borrower and due date for return of the material. This is necessary so that the library can recall an item, if needed, for another purpose (e.g. reserve) and to assure responsible use of material by patrons.

Current Methods

The Stanford University Libraries recorded over 870,000 circulation charges during the year 1968/1969. There is no central department in charge of all circulation procedures. Each circulation point makes its own regulations, keeps its own files, and maintains its own staff, although certain libraries exercise the option of sending unpaid bills for lost/unreturned material to the main library service desk for processing. Loan periods within the Stanford University

Libraries are: 2 hours, overnight, 24 hours, 48 hours, 72 hours, 4 days, 1 week, 2 weeks, 1 month, 1 quarter, and 1 year. In addition, some material is reserved for use in the library only and does not circulate. The libraries maintain information about items in circulation by cards filled out by the borrowers when they check out an item. Some libraries maintain parallel files of such cards in order to have accesses by call number, due date, and borrower. Libraries with large circulation volumes divide these files into general, reserve, faculty-staff and doctoral charge files to facilitate filing and to make it easier to read through the files to identify overdue items. There also may be additional files for requests for recall of material, fines, and books at the bindery. The libraries use a system of notices and fines to assure equitable availability of material to all users.

Limitations

1. COMPLEXITY OF FILE USE AND MAINTENANCE

Particularly in the libraries with a high circulation volume, file maintenance is a time consuming and expensive task which is prone to error. In order to provide all the needed access points and to simplify searching procedures, numerous files have been created. Because of the pressing need for coordinating numerous files, the librarians have in some cases, been forced to maintain procedures which are not entirely adequate. An example of this is the hold process. When a book is returned without the recall request, it may be placed on the shelf and checked out again before the library staff is able to identify it as an item for which there is a prior request.

2. USER INCONVENIENCE

The current charging system requires that the patron copy the call number, author, title of a book, and his name and address on a charge slip. He must repeat the process for each and every piece of material borrowed. This is an irritating process which is also error prone.

4.22.2 RESERVE PROCESSING

Objectives, products, and services

When a professor assigns a book as required reading for his class, a library places an appropriate number of copies in a special location and puts them on reserve circulation with a loan period of two-hours or one day. The preparation and record keeping associated with this process is termed Reserve Processing. While the books are on reserve, the library must provide some list or catalog of material so that they can be located and used.

Current Methods

The Stanford University Libraries placed over 45,000 volumes on reserve during the year 1968/1969. As is the case with circulation, each library does its own reserve processing. The libraries begin to receive reserve lists from professors at the end of each quarter for the next quarter's classes. The flow of lists is heavy through the break between quarters and on into the first part of the new quarter.

The first step in processing these lists is to see if the library has enough copies of each book to meet the expected demand. This requires a search of the library records and the shelves. If there are not enough copies, the library must obtain additional copies by loan or purchase. The library then makes a record for the general circulation file to indicate that these books are on reserve. The degree of control maintained over a reserve collection varies greatly from library to library. Some libraries maintain only a loose leaf binder listing the books by author, organized under the appropriate course numbers. Some maintain complete shelf list, author, and course files for all reserve materials. Often a library will place a book on 1-day reserve at the beginning of a quarter and shorten the loan period to 2 hours near the end of the quarter. In this case, all the records in the reserve files must be changed. At the end of the quarter the material is returned to the general circulation shelves unless the professor specifically requests that the material be retained on reserve for the next quarter.

Limitations

1. INABILITY TO MEET PEAK LOADS

The reserve processing work load is subject to considerable fluctuation throughout the quarter. It is an operation in which backlogs cannot be tolerated. If a book is not processed for reserve when needed, it is of no use. Much of the work comes at the end of the quarter, and during the break between quarters. This is a particularly difficult time because student help generally is not available during final examination and vacation periods. Therefore it is difficult to keep up with the necessary amount of manual file searching and typing of records.

2. MANUAL SEARCHING OF DATA ALREADY IN MACHINE READABLE FORM

The Meyer Undergraduate library processes more books for reserve use than any other library on campus. Despite the fact that bibliographic information for most of these books is in machine readable form, these same manual typing and searching procedures are employed.

4.22.3 REFERENCE SERVICE

Objectives, products, and services

The primary objective of reference service is to help the library patron fulfill his information needs, particularly by aiding him in his use of the library's collection. In a broad sense, reference service covers everything necessary to help the reader in his inquiries, including (1) the selection of an adequate and suitable collection of reference books, (2) the arrangement and maintenance of the collection in such a way that it can be used easily and conveniently, (3) the making of special files, indexes, lists, bulletins, etc., to help the reader find and use information, (4) instruction of individuals, groups, and classes in reference methods and the use of reference books, and (5) answering individual questions.

Reference Service - Current Methods

There are three areas in the Stanford University Libraries which are specifically organized and staffed to provide reference service. These are the Main Library General Reference Service, the Government Documents Department, and the Meyer Undergraduate Library. These units each have a number of professional reference librarians with specific assignments in certain subject areas. In the smaller libraries, the head librarian provides reference service in addition to his administrative responsibilities. In a few cases, where there are no professional librarians, reference service is provided by telephone to a larger library. All librarians involved in reference work spend a portion of their time: (1) answering reference questions, (2) doing research to stay current in their subject area, and (3) selecting books both for the reference collection and the general collections of the library.

Limitations

It is difficult to define limitations in the reference process because the elements and decisions are not fully known. Unlike many routine procedures, the reference process is not easily represented as a series of definite steps. For this and other reasons, reference needs to be studied to learn more about questions, the role of files and types of search strategy. Research will show potential uses for automated tools in the reference process. The value of the skilled reference librarian in this study is inestimable.

4.23 Summary of Limitations in the Current System

Four basic factors now limit the library in achieving its objectives:

1. DEPENDENCE ON MANUAL FILES

First, a record in a manual file provides only a single point of access. If there are to be multiple points of access, copies must be made of the record and these copies filed at each of the desired points of access. Second, the larger a file becomes, the more difficult and inefficient it is to find a given item in that file. It is much easier to find one specific item among ten records than one among ten thousand. Third, manual filing is subject to error, and these errors are difficult to locate after they have been made. In order to overcome limitations of access, duplicate copies of records are made and placed at various points in a file. But this makes a file larger, and therefore harder to search. To solve this problem, files are sometimes broken up into several smaller files. This trades the difficulty of searching a large file for the difficulty of coordinating many separate files.

File proliferation is also a result of the need for access to information at various library locations. Duplicate files must be maintained to meet these needs.

These factors degrade the quality and reliability of library files. This degradation is being retarded to some extent by an elaborate, time consuming, and expensive process of filing revision. However, the size and number of manual files has passed the point where even elaborate manual procedures can maintain quality without introducing other problems such as further error and higher cost.

2. INCREASING DIFFICULTY IN MEETING WORK LOADS

Current manual procedures are being pressed to their limits. In areas where work loads fluctuate, it is sometimes impossible to keep pace during periods of peak volume. As the volume of library processing increases, the capacity of manual procedures becomes saturated more frequently. Owing to the inefficiency inherent in large group operations, the addition of more personnel can not solve the problem. The point has already been reached where the addition of one employee yields a productivity net increase of less than one full employee.

3. UNUSED STAFF POTENTIAL

The current manual procedures cannot make efficient use of the library staff. They involve many menial clerical tasks such as typing, simple proofreading, filing, and file revising. These activities require accuracy, speed, consistency, and resistance to boredom. Such qualities are better asked of machines than of people. If the staff could be relieved of this type of work, they could devote more of their time to intellectual tasks requiring flexibility, logical discrimination, judgment, and imagination. This relief would allow the library assistants to take care of more complex activities and free librarians for work which requires their special training.

4. LACK OF CONTROL INFORMATION

Current manual procedures make it difficult to control the flow of material through the system. A considerable amount of redundant record keeping is necessary to provide basic information such as who has what and who ordered what. It can be difficult and time-consuming to locate a certain item or to obtain a report of its status during Technical Processing.

Lack of control information affects not only material, but also the allocation and management of library resources of all types. Thorough and comprehensive management statistics are needed in order to evaluate and improve current library procedures.

The consequences of these limitations are manifested in two ways: the degradation of service and rising costs. The Stanford Libraries have gone to great lengths to avoid degradation of service. In doing so, they have had to pay a higher price: rising costs. In an efficient operation, costs do not rise in direct proportion to volume. Unit costs should decrease as volumes increase. But in the expanding work load environment of the library community, the reverse has been true. For example, the unit cost of Technical Processing (e.g. the cost of preparing 1 book for a reader) at Stanford has risen almost 50% in the last five years.

The university is faced with rising costs, reduced funds in some areas and increased competition for available funds. The libraries of the university are called on to maintain a standard of excellence on a growing scale. The libraries of Stanford contribute to the overall character of the university both by the quality of their collections and the effectiveness of their operations. The most serious consequence of the limitations inherent in current methods is the decreasing ability of the libraries to make a maximum contribution to the educational quality of the university.

5.0 LONG RANGE SCOPE FOR LIBRARY AUTOMATION

This chapter addresses itself to those limitations described above that can be overcome in a cost-effective manner. It describes the functions and processes in the Stanford University Libraries in need of system support. The development of all facilities mentioned in the Long Range Scope will doubtless be spread over several development iterations. A subset of facilities will be isolated in Chapter 6 for development in the current iteration.

5.1 General Considerations

The long range system scope is an approach to the development of cost effective bibliographic processing for the university library.

The preliminary analysis phase just completed has established:

1. The library areas in need of computer support
2. The kinds of support required
3. The cost limits imposed on a production system
4. The growth capabilities required.

Experience with the prototype system has established that on-line bibliographic searching is applicable to a variety of library operations; that library automation requires the full cooperation of both the library and the university; That data preparation and control are critical functions for well managed coding, editing and input activities; and that library personnel can work effectively in a computer system environment.

5.11 Technical Processing

The long range manual-automated system will be be characterized by the following:

1. A system configuration with both manual and machine flexibility to accommodate wide fluctuations in input volumes.
2. A single In Process File, accessible on-line by multiple remote terminals, to control all items in process.
3. On-line search capability using multiple search access points against the In Process File to find (1) whether a record is in process, (2) the status of a record in process.
4. Use of nationally created machine readable bibliographic data when available. Particular attention will be given to the use of the Library of Congress MARC Distribution Service.

5. Ability to meet the requirements of libraries outside the Stanford University Libraries system.
6. The production of all required Acquisition and Cataloging printed outputs.
7. Automated accounting procedures for acquisition of library material.
8. Computer service for detailed control of serials holdings and acquisition/binding activity.
9. Facilities for management reports for both manual and automated processes.
10. Computer service for Binding and Finishing control.
11. Automatic material and invoice claim control.

5.12 User Services

5.12.1 Circulation

The Long Range manual-automated system will support all circulation functions at all library service points during regular library service hours. The circulation functions to be supported are:

1. charging - including:
    - Both cataloged and uncataloged materials
    - all of the various circulation periods
    - ability to add, change, or remove any circulation period
2. discharging
3. overdue material processing - including:
    - identifying overdue circulation charges
    - notifying borrower holding overdue material
    - calculating fines, with attendant record keeping
    - notifying borrower about outstanding fines
4. billing for lost/not returned books - including:
    - identifying unreturned material
    - calculating bills, with attendant record keeping
    - notifying borrower
5. collecting delinquent bills - including:
    - identifying delinquent bills
    - notifying both borrower and registrar
    - record keeping
6. processing of requests for holds on material in circulation - including:
    - recording hold requests
    - identifying returned books with hold requests against them
    - notifying requester when material is available

7. recalling of material in circulation - including:
    - recording recall requests
    - notifying borrower to return material
    - notifying requester when material is available
8. renewing of charges - including:
    - quarterly doctoral charges
    - annual faculty/staff charges
9. searching of the circulation file to determine if a book is in circulation
10. statistical record keeping and analysis

Shelving of books and searching of library stacks will, of course, remain manual procedures, but will be interfaced with the automated portions of the system.

Essential characteristics of the circulation system will be:
1. ability to handle both cataloged and uncataloged material
2. automatic self service charging by the patron
3. automatic discharging
4. automatic recognition of returned material for which there is a hold request
5. on-line searching of the circulation file
6. machine readable book identification
7. machine readable borrower identification

5.12.2 Reserve Processing

The library automation system will support reserve processing. It will be able to supply special support to libraries that have machine readable data bases available. Reserve Processing functions to be supported are:
1. searching of machine readable data bases
2. production of purchase orders
3. production of reserve book processing slips
4. production of book identification for reserve circulation
5. charging of books from regular circulation to reserve
6. production of reserve catalogs
7. statistical record keeping and analysis

Essential characteristcs of the reserve processing system will be:
1. ability to handle both cataloged and uncataloged material
2. on-line searching of available data bases
3. ability to accomodate peak loads

5.12.3 Other User Services

The areas of Inter-Library Loan, Technical Information Service, and Reference Service are also within the scope of the automated library system. It is not possible at this time to determine what specific functions in these areas are in need of, and amenable to, computer support. The research and analysis necessary to determine the type of support appropriate for these areas is within the scope of the continuing library automation effort.

6.0 FIRST IMPLEMENTATION SCOPE

As mentioned in the discussion of the Long Range System Scope, an integrated system, servicing both Technical Processing and User Services, will be the focal point of the next development cycle.

The system scope for Technical Processing is represented primarily in flowchart and narrative form. The facilities for User Services are described in narrative form.

6.1 Technical Processing

I. General Features

The Technical Processing system will service both the Acquisition and Catalog Departments of the Stanford University Libraries. The general features of the system are:

1. One time capture of bibliographic and control data during acquisition processing for: data preparation, input, record creation and subsequent generation of required outputs.

2. File updating as the status of an item in process changes. For example, material receipt, order cancellation, or cataloging completed.

3. The production of all major outputs as the result of updating. For example, replacement purchase orders, claim notices, cancellation notices.

4. Use of MARC bibliographic data for acquisition and cataloging outputs.

5. The use of a single record to satisfy searching and output requirements as they arise in the Technical Processing cycle. For example, book circulation

identification, call number, spine labels, and catalog cards.

6. Management statistics and special activity reports. For example, vendor performance reports.

II. Detailed Description -- Acquisition

The Technical Processing system will support the searching, material purchasing and receipt, and end processing activities of the Acquisition Department. The system will include automated support for material and invoice claiming; invoice payment; computer produced spine label information and book circulation identification; the use of MARC bibliographic data; and the automatic collection of management statistics and special reports.

The primary acquisition file will be an on line In Process File.

The acquisition system is graphically represented in the following flowchart. A detailed description of the sytem inputs, manual processes, computer processes and outputs follows the flowchart. Call number spine label information and book circulation identifications will be automatically produced after the Catalog Department has updated the In Process File. These outputs will be used by the Binding and Finishing Division of the Acquisition Department for their end processing requirements. Binding and Finishing is not represented in a separate flowchart.

A. Inputs to the manual-automated system

Inputs to the system are of three basic types

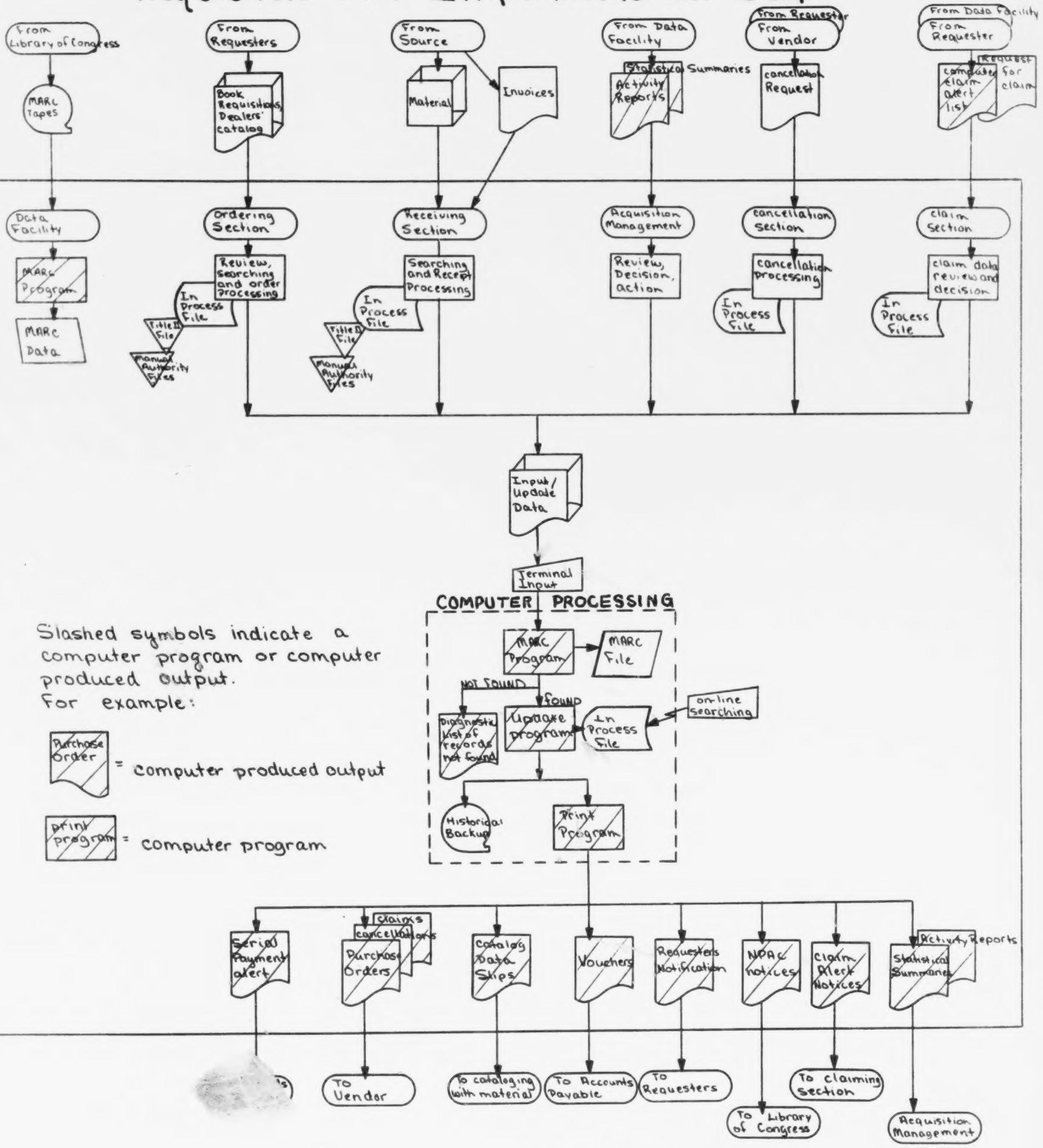
1. Communications from vendors, requesters, and other system users.
The Acquisition Department receives notices from a variety of sources requesting a specific type of action, such as a purchase request or a request to claim material.

2. Material and Invoices.
The Acquisition Department receives library materials from gift donors, exchange partners and vendors. Invoices are received from vendors.

3. Computer produced inputs to the acquisition system.
The acquisition system will automatically generate special listings and reports for management action, such as serial payment and claim alert listings and vendor performance reports.

Stanford University Libraries

Acquisition-First Implementation Scope


From Library of Congress
MARC Tapes
From Requesters
Book Requisitions, Dealers' catalog
From Source
Material
Invoices
From Data Facility
Statistical Summaries
Activity Reports
From Requester
From Vendor
cancellation Request
From Data Facility
From Requester
computed claim alert list
Request for claim
Data Facility
MARC Program
MARC Data
Ordering Section
Review, Searching and order Processing
In Process File
Title II File
Manual Authority Files
Receiving Section
Searching and Receipt Processing
In Process File
Title II File
Manual Authority Files
Acquisition Management
Review, Decision, action
cancellation section
cancellation processing
In Process File
claim Section
claim data review and decision
In Process File
Input/ Update Data
Terminal Input
COMPUTER PROCESSING
MARC Program
MARC File
NOT FOUND
FOUND
Diagnostic list of records not found
Update program
In Process File
on-line searching
Historical Backup
Print Program
Purchase Order
= computer produced output
print program
= computer program
Serial Payment alert
Purchase Orders
cancellations
claims
Catalog Data Slips
Vouchers
Requesters Notification
NPAC notices
Claim Alert Notices
Statistical Summaries
Activity Reports
To Vendor
To cataloging with material
To Accounts Payable
To Requesters
To Library of Congress
To claiming section
Acquisition Management

Slashed symbols indicate a computer program or computer produced output.
For example:

EAM 1/27/70

B. Manual processes

1. Searching
Includes human decisions; its path varies with the type of item being searched. Searching involves, in summary, checking existing manual files, computer files, and printed books and catalogs to determine the status of an item and to determine the appropriate action.

2. Acquisition processing
Involves the manual procedures necessary to act on a communication coming into the Acquisition Department. Examples include claiming, cancelling, end processing and material and invoice receipt processing.

3. In Process File update processing
Involves the preparation of updates for transactions and decisions concerning an In Process File record.

C. Computer processing

1. MARC
Marc bibliographic data from the Library of Congress will be used for acquisition and cataloging processing. Methods for the processing of MARC tapes, the extraction and conversion of MARC records and the use of the MARC data will be determined during the Detailed Analysis Phase of System Development.

2. In Process Record Creation and Update
New acquisition and brief bibliographic information and all subsequent updates will enter the In Process File. Outputs and special activity and statistical reports will be generated as required for subsequent printing and formatting. Where required, computer produced outputs will be sorted according to predetermined criteria. Historical data will be kept after all Technical Processing is completed for a given record.

The In Process File will be accessible by several points (for example, record identification number, author, title) and will be available for on line searching.

D. Computer produced outputs from the manual-automated system

1. Purchase Orders, Cancellations and Claims. Used to communicate information to vendors.

2. Catalog Data Slips. Used to communicate bibliographic and acquisition data to the Catalog Department with the material.

3. Vouchers. Used to communicate fund and billing data to the University's Accounts Payable unit

4. Requester Notification: Used to communicate information about a requested item to its requester.

5. National Program for Acquisition and Cataloging (NPAC) Program notices. Used to report acquisition and bibliographic information to the Library of Congress as part of the NPAC program.

6. Statistical Summaries and Special Activity reports such as serial payment and claim alert. Used to communicate to staff and management processing statistics and special activity reports

III. Detailed Description -- Cataloging

The Technical Processing system will support catalog card production for the Catalog Department. As the following flowchart indicates, initial consideration will be given to computer produced catalog cards for items with MARC bibliographic data. As a result of the Detailed Analysis Phase, this scope may be expanded to include the use of other sources of bibliographic data for computer produced catalog cards.

A. Inputs to the manual-automated system

1. Includes library material received in the Catalog Department.

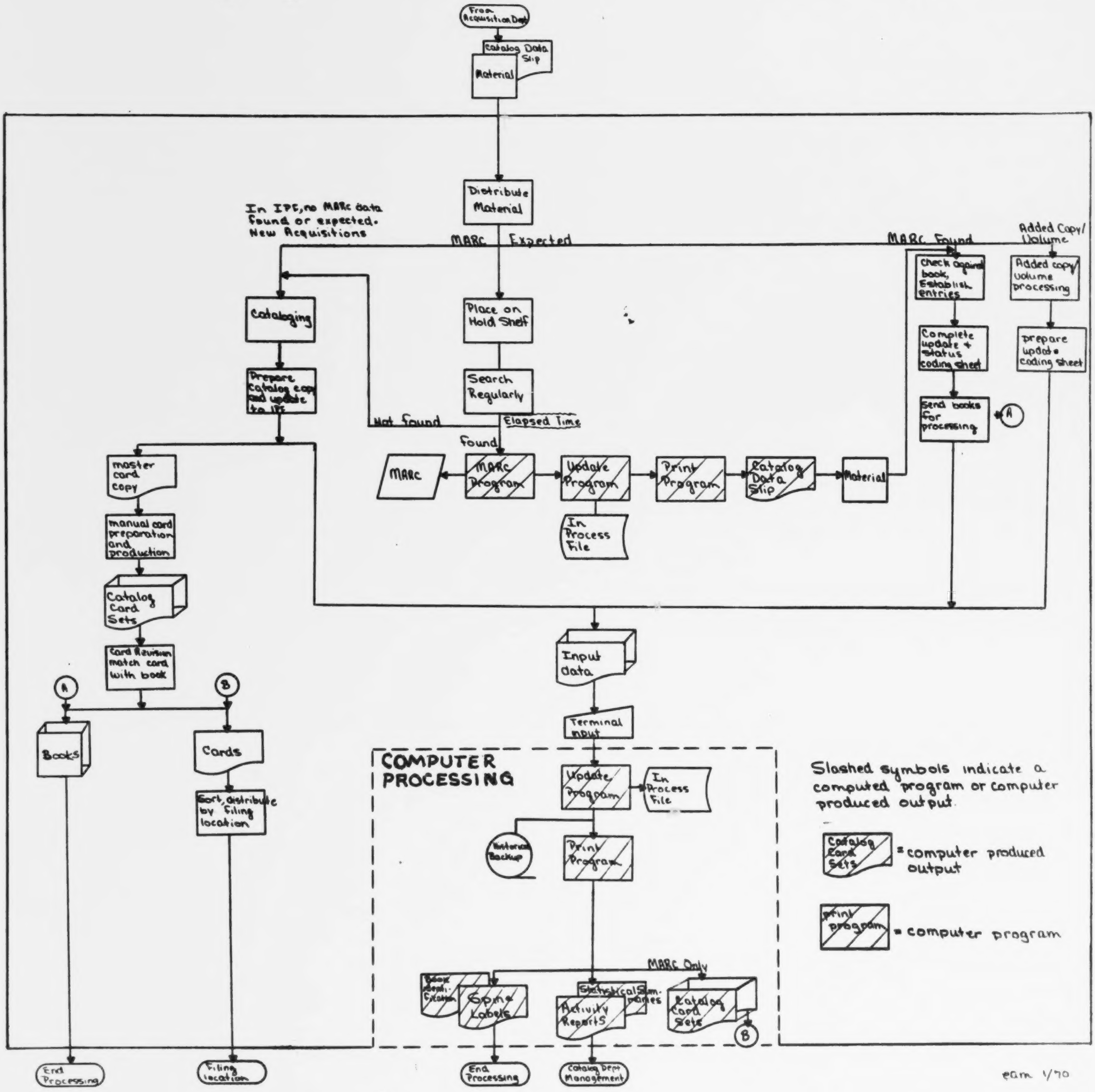
2. Catalog Data Slips Used to accompany material to the Catalog Department to communicate pertinent control, bibliographic and special message information found during acquisition processing.

B. Manual Processes

Catalog processing involves many intricate procedures which vary according to the type of item being processed. In summary, cataloging processing involves:

1. Searching manual and computer files to find Library of Congress bibliographic information and information about items already held.
2. Creating a full bibliographic record for an item for which no bibliographic record exists.
3. Maintaining already existing files and records. For example, adding volumes and copies, transferring and cancelling volumes, changing call numbers.
4. Modifying and using existing bibliographic data to create a bibliographic record. For example, LC Card cataloging.
5. Card preparation and production for material not in the scope of the Automated System.

Stanford University Libraries
Cataloging - First Implementation Scope


From Acquisition Dept
Catalog Data Slip
Material
Distribute Material
In IPF, no MARC data found or expected. New Acquisitions
MARC Expected
MARC Found
Added Copy/Volume
Cataloging
Place on Hold Shelf
Check against book, Establish entries
Added copy/volume processing
Complete update + status coding sheet
prepare update coding sheet
Prepare catalog copy and update to IPF
Search Regularly
Send books for processing
A
Not Found
Elapsed Time
Found
master card copy
MARC
MARC Program
Update Program
Print Program
Catalog Data Slip
Material
In Process File
manual card preparation and production
Catalog Card Sets
Input data
Card Revision match card with book
A
B
Terminal input
Books
Cards
COMPUTER PROCESSING
Update Program
In Process File
Sort, distribute by filing location
Historical Backup
Print Program
MARC Only
Book Identi-fication
Spine Labels
Statistical Summaries
Activity Reports
Catalog Card Sets
B
End Processing
Filing location
End Processing
Catalog Dept Management
Slashed symbols indicate a computed program or computer produced output.
Catalog Card Sets
= computer produced output
Print program
= computer program

eam 1/70

6. Filing in manual files.
7. Preparation of updates to In Process File.

C. Computer Processing -- see Acquisition Computer Processing

D. Computer produced outputs from the manual-automated system

1. Statistical Summaries and Activity Reports. Used to communicate statistical and activity data to staff and management

2. Spine label information and Machine Readable Book Circulation Identification. Used by Binding and Finishing Division for material preparation for Circulation (End Processing).

3. Catalog Card Sets. Used to file in Stanford University Libraries manual files to indicate an item is held by the library. Catalog cards will be produced from MARC data. Further study is needed to determine how much of the Catalog Department non-MARC output will be captured for catalog card production and future use.

IV. Areas in need of additional study

The first implementation scope for Technical Processing has been chosen with the knowledge that several areas are in need of additional study. The result of study in these areas will affect the first implementation.

The following areas will be examined in detail:

1. Exchange. Feasibility of servicing the processing requirements of the Exchange Division.

2. Capture of cataloging data. Economics of manual card preparation and production for material not included in MARC must be compared with the costs of capturing the data in machine readable form for subsequent automated catalog card production.

3. Book pocket labels. The economics of machine produced book pocket labels

4. Machine readable accounting data. The feasibility and economics of creating machine readable accounting data as input to the University Controller's Accounting system

5. Machine readable authority files. The economics of creating and maintaining machine readable cataloging authority files

6. Selective dissemination of information and special reports. The economics of using In Process File data to prepare automatically SDI lists and special reports, such as recent acquisition lists

7. On-line Science Union Catalog. Study of a small, well defined subset of the present manually maintained Science Union Catalog

6.2 User Services

The First Implementation scope in the area of User Services will encompass the circulation and reserve processing functions of the J. Henry Meyer Memorial Library (undergraduate library).

6.2.1 Circulation

The first implementation scope will service the Meyer circulation system as a whole. It will provide computer support for the following processes:

1. charging of all circulating library material (reserve and general)
2. discharging of all circulating library material
3. collecting fines
4. billing for lost/not returned books
5. collecting delinquent bills
6. processing hold and recall requests
7. searching circulation files
8. statistical recording keeping and analysis

The following procedures will be integrated into the circulation system as manual procedures:

1. shelving books
2. searching shelves for lost books

The aim in the first implementation is to produce an on-line, self-service circulation system using machine readable book and borrower identification. The existence of equipment and technology to support such a system reliably and at a reasonable cost is currently open to some question. After the detailed requirements for the system have been defined, it will be possible to answer this question definitely. The matter of machine readable identification for library patrons is also to be resolved. These two factors may affect the implementation of the circulation system.

6.2.2 Reserve Processing

The first implementation scope will service the entire Meyer Reserve Processing system, including:

1. searching the machine readable Meyer cataloging data base
2. ordering of material needed for reserve which is not held by the library
3. production of processing slips used in preparing books for reserve
4. production of reserve catalogs
5. statistical record keeping and analysis

The aim in the first implementation is to produce a reserve processing system with on-line searching for reserve material and computer production of all outputs from reserve processing in order to provide the library with faster, more efficient service in this area.

PART III

GENERALIZED INFORMATION STORAGE AND RETRIEVAL

7.0 CURRENT STATUS, GENERALIZED INFORMATION STORAGE AND RETRIEVAL

The SPIRES I Generalized Information Storage and Retrieval (GISR) Facility has been operating as a prototype System for approximately one year. During that time, the Stanford University Libraries, the Stanford Linear Accelerator Library, the ERIC Clearinghouse, the Department of History, and the Department of Geology have all built, maintained, and searched files on-line. Thus, it is seen that users of this facility do not fall into any particular organizational hierarchy, but are widely distributed geographically and with respect to academic discipline. Furthermore, the system now in existence and any system yet to be designed in no way changes the user organization or his procedures beyond those used for information gathering. These two facts make it necessary to weight the GISR discussion of current operations heavily toward software facilities as opposed to organizational divisions, functions, and processes.

7.1 Representative User Profiles

Various types of bibliographic users could easily make use of a GISR capability. There follows a brief sketch of seven possible user types. Refer to appendices E and F for detailed descriptions of law and physics users.

DEPARTMENTAL LIBRARIAN

Librarian Smith in a departmental library has been following the literature on machine-assisted bibliographic searching. A number of department members have made inquiries regarding a subscription service for computer tapes containing comprehensive bibliographic information in their field of interest. Librarian Smith does not know anything about computers but she is willing to learn in order to get a copy of the data collection. She does not do bibliographic searching for members of the department at the present time. In the future she would be willing to search the data collection for those professors who did not want to learn how to use the computer. Librarian Smith does not have any assistants.

RESEARCH LIBRARIAN

Librarian Brown of the university professional school library is an outstanding researcher. His library staff does most of the bibliographic searching for the faculty of the school, and occasionally for outsiders. He has determined that a considerable amount of searching time could be saved if the literature in an emerging field were properly indexed and kept up to date. He realizes that his school cannot afford to do this work in isolation, and so proposes to serve as a clearinghouse for indexing in the field. He is skeptical of computers but sees no manual method for preparing the material and keeping it updated without a large staff.

SENIOR RESEARCHER

Professor Black is a tenured member of the department and has an international reputation. He is a prolific writer and is the senior member of several research teams. Because of his heavy workload, he cannot afford to do bibliographic research personally. He hires graduate students to do the work, but is discouraged by the uneven quality of their work. If a device could be provided to allow him to search existing files exhaustively and rapidly, he could find what he needs more efficiently and use the graduate students for more exciting work.

EXPERIENCED RESEARCHER

Professor Lang has a collection of data relating to California. In his collection he has public opinion survey results, election returns, and census data. He wants to store this information on-line in card image format so that he and his students can test a series of behavioral hypotheses. Instead of listing the data resulting from a search (except for frequency counts, display of questionnaires, or candidate names) it would be saved for use by statistical routines.

INEXPERIENCED RESEARCHER

Instructor Jones is young and new to the department. He usually works alone because most of his colleagues do not work at the same pace. There is no adequate index to research literature in his specialty. Because of his experience with computers as a student, he wants to build a bibliographic data collection. He proceeds to build the collection and uses it extensively. After a year of work during which a 500 document collection is accumulated, his interest turns to a different problem in a related field. He moves to another university and his collection is abandoned.

RESEARCH ASSISTANT

Graduate student Johnson is a heavy user of the departmental library. He feels that he spends too much time trying to find material relevant to his interests. Since he has had experience with computers as an undergraduate, he considers it obvious that computers could be used to assist him. However he is afraid to rely too heavily on the computer since other universities might not provide the same services.

VISITING RESEARCHER

Mr. Peters is a graduate of the university but is now working in industry. He often needs to do research in his field. He feels uncomfortable when he visits the departmental library because he does not know anyone and does not know how the material is organized. He does not know much about computers and would use one only if led by the hand. He is willing to pay to get the help he needs.

7.2 Summary of User Requirements

The needs of the users profiled above form a wide spectrum. The requirements of Librarian Smith are complex and involve many capabilities for which library funds might be available; the graduate student has a well defined problem and at best a small budget to expend in solving it. Most other users fall somewhere between these two extremes.

ECONOMY & EFFICIENCY

The system must have a file structure that optimizes the trade-off between response time and disk storage utilization. Furthermore, the system software must be as efficient as possible while the hardware configuration must have just enough capability to do the job and no more. The cost for terminal time and for storage of information on-line must be low enough to be attractive.

SIMPLICITY

A successful system is usually simple to use. Some users have no computer background, and others have experience of relatively short duration. It is therefore necessary that a beginner be able to acquire the knowledge he needs with a minimum of research and study, preferably by having the system "lead him by the hand" during the initial phases. Furthermore, when the user commits an error, he should be directed toward resolution of his problem by a carefully conceived set of diagnostic messages.

FLEXIBILITY

The successful system must be user-adaptive, providing a variety of facilities to satisfy every need and pocketbook. A sophisticated system is obviously costly; if a simple and basic capability will suffice, the user should be given just that and charged accordingly. A consequence of this flexibility is that each user's file will look different. Thus, the need for AUTOMATED FILE DEFINITION (see 7.31.2 below) presents itself.

FEEDBACK

In order to evaluate the performance of the system, it is necessary to gather statistics which show the nature of the data stored in the system, the means used to retrieve it, and the frequency of access. Given such information, users may re-evaluate their file content and definitions in light of their experience, and make changes where appropriate. In addition, feedback must be provided regarding frequency of use (by user type and file type) and frequency of errors committed by users or by the system.

7.3 Summary of Current Facilities and Limitations

This summary of SPIRES I current facilities and limitations will entail brief descriptions of the two portions of the prototype system: data management and retrieval. Data management refers to the preparation, collection, formatting, storage, and maintenance of bibliographic information. Retrieval refers to the use of this information by people with the aid of the SPIRES/BALLOTS system. Both portions of the system are based on a file structure designed to provide maximum flexiblility in the placement and retrieval of data.

7.31 Data Management

Data management under SPIRES I refers to the manual-automated facility designed to handle data preparation, the establishing of files, file maintenance, and any special applications.

7.31.1 Data Preparation

The input of data into the system by local keyboarding and by conversion of data already in machine-readable form are the two means of data preparation. In either case, the end product is data in SPIRES Update Command Language format which is acceptable to the file building and updating program.

INPUT OF RAW DATA

The gathering of raw data is achieved by clerical workers using WYLBUR, the Stanford text editing facility. This method is more flexible for many applications than the alternative of keypunching card decks to be read into the system.

CONVERSION OF MACHINE-READABLE DATA

Large quantities of bibliographic data are available in machine-readable format. Such data is received on magnetic tapes which can easily be mailed from anywhere in the world. Conversion programs have been written to make some of these formats acceptable to the SPIRES system. DESY and NSA tapes (high-energy physics) can now be converted, as well as ERIC tapes (Education Research) and MARC (Library of Congress Machine-Readable Catalog).

SPIRES UPDATE COMMAND LANGUAGE FORMAT

The SPIRES Update Command Language format was designed for ease of encoding by human beings. It has, therefore, served its purpose adequately for data keyboarded locally. However, as a format into which to convert machine-readable data, the Update format has meant unnecessary inefficiency. A highly compact intermediate format into which to convert both SPIRES Update Command Language data, and other machine-readable formats is needed. Such an intermediate format would alleviate the decoding of highly compact machine-readable data into human-efficient format, which then has to be immediately re-encoded in the SPIRES files. Regardless of this drawback, the conversion process was a valuable feature of the SPIRES I system.

7.31.2 Establishment of Files

Prior to any file building or updating, files are defined and established. System programmers and users together determine how much disk space is required, the data elements to be used, data element values to be expected (format, length, multiplicity), which ones are to be indexed, and any special editing to be done. File definition under SPIRES I is done manually, and programmer assistance is required. An automated system was developed to interpret commands in a File Characteristics Language and generate a user-specific file definition, but it was not interfaced with the rest of the system. The next SPIRES system, in addition to automating the definition of these parameters, should look to other areas of user specification. For example, the definition of a large storage/low usage file might be distinguished from that of a small storage/high usage file, in such a way that efficiency and performance could be optimized in either case. This implies that the results of such file definition would be utilized by all parts of the system, not just by the data management portion.

7.31.3 File Maintenance

File maintenance under SPIRES I is accomplished by means of a batch mode record level Update facility. That is, one can add entries to the file and delete them, thereby replacing any entry. The use of storage in this task was geared toward reclamation of unused disk space. Therefore a dynamic file (heavily updated) would not grow indefinitely, but reach a point of space utilization equilibrium. In addition, statistics are kept regarding numbers of entries and data elements, and regarding questions of space and structure. Bibliographic entries are restricted in length to about 3500 characters of information and file size is limited by hardware capacity.

Various file management aids were developed to ease the task of the non-technical data manager. In particular, an experimental on-line macro facility was developed to aid the manager in such tasks as initiating build and update runs on the files, maintaining backup copies of those files on tape, and restoring files when necessary. This allowed the file manager to proceed somewhat independently from the system programmer in the file maintenance task. Further steps in this direction will be taken in future SPIRES/BALLOTS systems.

7.31.4 Special Applications

The development of any automated system involving files and useful information often encourages special applications not envisioned in the original system design. SPIRES I has been no exception. Data prepared for input to the system has also been used to produce PREPRINTS IN PARTICLES AND FIELDS, a weekly newsletter containing the most important bibliographic information sorted by key. In addition, the SPIRES data base has been used to produce for SLAC a semiannual publication containing bibliographic descriptions of articles by local authors only, sorted by author, subject, and key.

7.32 Retrieval Facility

The process by which bibliographic data is entered into the system and kept current has been discussed. What follows is an explanation of the means by which data is retrieved.

The SPIRES Retrieval system is a fully automated on-line bibliographic search capability allowing the remote terminal user to make various search and output requests.

7.32.1 Search Requests

Once communication is established with the retrieval facility, the SPIRES user must select a specific file for bibliographic searching. For example, he might choose the SLAC Preprint file or the Geology file. The user may then begin an interactive search session on his selected file. Depending on his choice, he may search on such indexes as are available for that file. Author indexes can be searched on names in a variety of conventional formats (first last; last, first; etc.). Titles are searched by specification of one or more title words or title word stems which do not appear on the system exclusion list (words too heavily used to be meaningful as search items). Citations require a more rigid format: journal description, volume number, page number. The user may interactively narrow or broaden his search by compound search requests, using the connectives AND, OR, and NOT to combine search terms from any index. Search results may be further narrowed by specification of dates: BEFORE, AFTER, FROM, SINCE, or THRU may be used. If the searcher finds he has inadvertently narrowed his results too far, he may BACKUP to his earlier findings.

7.32.2 Output Requests

At any point in the search session, the user may interrupt his searching and have his accumulated results typed at his terminal. He may use the standard SPIRES output format, which includes all data elements in each document and their associated values. Or, he may select certain data elements to be listed in a specific order. In using this second option, the user could have the title printed first, and if it were of interest to him, allow the rest of the document description to be printed out, otherwise interrupting the output and going on to type the next entry.

7.32.3 General Comments

A SPIRES Reference Manual has been published which contains a step by step description on the use of the SPIRES Retrieval Facility. It would have been desirable to have incorporated more of this training into the system itself in order to ease the user-initiation process. This would imply a more extensive error diagnostic and error recovery capability. In terms of output of search requests, a print off-line capability is certainly needed. Another feature needed in a future version of SPIRES/BALLOTS is the manual and automated use of statistics on the retrieval facility to improve overall system performance, efficiency, and responsiveness to users.

8.0 Long Range Scope for Generalized Information Storage and Retrieval

The preceding section dealt with the present system, SPIRES I. This section defines those facilities to be eventually added to the system. It must be noted that some, but not all, will be chosen as a Scope for Implementation in the next development iteration.

8.1 Retrieval

Retrieval requests will have two essential parts: a search request and an output request. A series of iterative search requests, each giving feedback to allow framing of subsequent requests, will state the criteria which the user wishes any retrieved record to meet. The output request will state which data elements of the retrieved records he wishes to see. These facilities will be available for both on-line and batch operations.

8.11 Search Requests

INDEX TYPES

The basis for on-line retrieval is the set of indexes associated with a file. There exist many kinds of indexes; each index represents a different way to enter the file. Some examples are given below.

1. Personal name indexes: Personal names consist of alphanumeric characters. Names are indexed with surname first, followed by given names (or initials), followed by title, if any. For example, the name "Sir John Gielgud" would be indexed as "Gielgud,John,Sir". In retrieval, this allows matching on phonetic representations, surnames only, surnames and initials, or an exact match on the full name, e.g., FIND EMPLOYEE MOOK, EMPLOYEE MOEK, or EMPLOYEE L. MOEK, or EMPLOYEE LARRY J. MOEK. The more specific the request for a match, the fewer matches are found.

2. Title word index: Titles consist of one or more words comprised of alphanumeric characters. Each significant word in the title phrase is indexed separately. In retrieval, a match on a single word will retrieve all titles containing that word. A match on a word phrase could result in retrieving all titles containing all the words in the phrase regardless of order, e.g., FIND TITLE HONEY BADGER would retrieve the titles: THE HONEY BADGER and THE BADGER WHO LIKED HONEY. Alternatively, specification of a

word phrase could result in retrieving titles containing an exact match, e.g. FIND EXACT TITLE HONEY BADGER would retrieve only the title: THE HONEY BADGER.

3. Topic index: Topics, keywords or subjects are all synonymous with the concept of specifying words and phrases which describe the subject matter treated in a document. Topics consist of one or more words comprised of alphanumeric characters. The entire phrase is indexed as a whole, not separated into individual words as with titles. In retrieval, the exact word or phrase is matched with order preserved.

4. Numerical indexes: Numerical indexes contain data element values comprised of integer characters. Each data element value is indexed once, e.g., numbers assigned to parts in a garage supply warehouse. Another type of numeric index would enable users to retrieve from a range of numeric values rather than only one specific value.

5. Date indexes: Since dates may be entered in various formats, they will be converted to a standard format before they are indexed. Examples of dates are: DATE OF PUBLICATION, DATE ADDED TO FILE, etc.

6. Coded indexes: Codes are comprised of alphanumeric characters. The code value is indexed once and matches for retrieval are made on the complete value. Dictionaries are used to convert the codes to their full equivalent. An example is a large manufacturing concern with many outlets across the country. Each outlet is assigned a code so as not to maintain the full name of the outlet in the indexes.

7. Broad classification on indexes: Some document collections can be broken into a few broad classes. When it is desired to index that kind of data, special consideration must be given to the fact that all the data falls into just a few groups. An example can be drawn from the SLAC Preprint files where all documents can be classified as containing experimental, theoretical or instrumentation information. It is desirable to be able to access the files of data through this classification, e.g.,all documents by Jones in experimental physics.

The above examples do not comprise an exhaustive list. Most data elements to be indexed can be classified into one of these categories. Facilities will nonetheless exist for defining those that do not.

MULTIPLE LEVEL ACCESS

In addition to the ability to define multiple access points for a file, users will have the ability to divide a file into several levels. Indexed elements will be used to select a set of records from a file. This set may be further searched using set of indexed elements or may be searched sequentially to check non-indexed elements against another set of criteria. For example, a search might be performed on a set of insurance policy files for all policies of a particular type issued during the year for a face amount of $5,000 or more. In this example, the access points would be the policy type and date. The sequential search would be performed on the amount.

BATCH SEARCH

An alternative to on-line retrieval will be batch retrieval. Batch requests may be formatted on-line, and syntax checked for correctness of structure. They will then be accumulated for later processing against the desired file. The file will be searched sequentially for matches of requests with stored information. To minimize repeated passes over the same items, the requests may be grouped so as to find all requested information from the first record before moving on to the next.

Batch retrieval restricts the way one formulates a search request. A user will not have the ability to expand or contract a set of selected items resulting from a single batch search. Several more batch searches may be required before the user finally retrieves the desired set of documents. In contrast, the manner in which one formulates a query for on-line retrieval of information is dependent upon the ability to access that information directly without passing over previously stored information. One can skip back and forth within the file gathering information, expanding or contracting the set of selected items, and examine the contents of that set when desired -- all during one session at the terminal.

SIMPLE SEARCH REQUESTS

In stating a query, the user will indicate which element or elements he wishes to access, e.g., AUTHOR. He will then supply a value against which all values for that particular element are compared, e.g., AUTHOR JOHN BROWN. Such a query would be a "simple request".

COMPOUND SEARCH REQUESTS

A facility will be available to construct compound requests. Simple requests may be combined into a logical expression by using the words "and", "or" and "not". The use of "and" will allow the user to specify two or more criteria which all the records retrieved must satisfy, e.g., AUTHOR BROWN AND TOPIC NUCLEAR PHYSICS. Using "or" will allow the user to specify two or more criteria, at least one of which must be satisfied in each record retrieved, e.g., AUTHOR BROWN OR AUTHOR JONES. The use of "not" will allow the user to specify a term which is to be excluded from the set of retrieved records, e.g., AUTHOR BROWN AND NOT AUTHOR JONES.

In addition to the logical expression capability, one will be able to group simple or compound requests so as to imply logical preference or ordering, e.g., (AUTHOR BROWN OR AUTHOR JONES) AND TOPIC NUCLEAR PHYSICS. In this example, parentheses are used to indicate a preferred grouping. Everything within the parentheses would be evaluated prior to performing the remainder of the request. One would be able to nest these groupings as in (AUTHOR BROWN AND ((AUTHOR JONES OR AUTHOR SMITH))) AND TOPIC NUCLEAR PHYSICS.

In response to a request, the system will indicate to the user the number of items found in the specified file for each simple request. If the request was formulated as a logical expression, the system will respond with the number of records that satisfy the complete request. The user now has several options. He may choose to browse through the content of the records, i.e., make a request of the output facility described later in this section. He may choose to begin a new search request on the same file or on another file. Or, he may wish to modify the previous request. By modifying the request, the user would expand or contract the set of retrieved records. For example, the request:

FIND AUTHOR JONES OR AUTHOR BROWN

might retrieve 75 records which have either JONES or BROWN as an author. The user might then enter:

AND TOPIC NUCLEAR PHYSICS

which will reduce the set to those documents which have NUCLEAR PHYSICS specified as a topic. The user may find he has narrowed his search too far and may then choose to use an OR to expand the set. If at any time the user finds he has made a poor choice of criteria, he will be able to return to some previous point in his query and start again from that point.

A search request may be qualified with a date. A search may be limited to only those items before or after a specific date or within a range of dates. This facility will allow a user to search through current information, i.e., that portion of a file added since some date. Other dates that could be used in this way are publication date, date added to file, etc.

WEIGHTED SEARCH REQUESTS

The search facility, as it has been described so far, is a "hit or miss" process. Either all criteria are satisfied for a specific record or nothing is retrieved. One may therefore wish to attach percentages or weights to the search terms in a request. Through the use of this facility, he will specify that all items be found which contain a specific number of a given set of terms, e.g., find all documents which contain any three out of five given terms. Another way of attaching weights to particular terms would be to submit a request for all records found exceeding a specified score, when each term is assigned a weight. For example, the following request:

> FIND TITLES (METAPHYSICS, 7 EPISTEMOLOGY, 5
> ONTOLOGY, 5 EXISTENCE, 4 PHILOSOPHY, 3) WITH
> TOTAL SCORE 9

states that all documents are to be found having titles with a combination of the words in parentheses, such that the sum of the attached numbers is nine or greater. Thus, the bibliographic items for the titles "Epistemology as a Philosophical system" and "Epistemology and Ontology" would be retrieved, whereas those with title "Existence - a Philosophical Examination" or "a Philosophical Examination of History" would not. This facility is generally called weighted searching.

An alternative scheme would provide for the specification of weights in terms of decimal numbers less than one, with search results ordered by descending score.

CORRELATION OF SEARCH REQUESTS TO ABSTRACTS

If a bibliographic file had a data element which contained abstracts, a retrieval criterion could be stated in terms of one or more English sentences. The retrieval process would correlate the given phrase with each abstract and retrieve those records containing abstracts with a correlation coefficient greater than some specified value.

It should be noted that Salton et al. at Cornell University have been experimenting with this facility for some time, but have not implemented an economical system. Such a facility lies beyond the current economic boundaries for SPIRES II.

DICTIONARIES

Dictionaries will be available to assist the user in selecting search terms. Some dictionaries may be general and applicable to all files while others may be specific to a particular set of related files. Dictionaries containing exclusion words, synonyms, codes and abbreviations would be specific to a set of related files. Dictionaries of this type will be built at the time a file is established and relate to the content material. The user will have the option to modify basic lists provided by the system to meet his own requirements.

A user may use synonym, and abbreviation dictionaries to guide him toward a selection of terms which are appropriate for the particular file from which he is retrieving information. A file may contain abbreviations unfamiliar to the user. He may be using a meaningful word or phrase in his request, but the file manager preferred and used another word or phrase in his indexing.

Some information may be stored in a file in coded form to conserve space. A dictionary is needed to find the full equivalent which the codes represent, e.g., scientific journal names maintained as coded data in the file with a dictionary giving the full names of the journals and their associated codes.

For other elements of a file, there are values or words which either have no significance as far as content is concerned or occur too frequently to be of much value in retrieval. For such elements, a file manager may construct a dictionary called an exclusion word list. Words on this list would be dropped from any request which included them. The user will have the facility to interrogate these lists.

THESAURUS FACILITY

The thesaurus facility will be closely related to dictionaries. A thesaurus is file-specific and may contain a list of synonyms for key words or phrases used in a file. Reference to this list will enable the user to select other words and phrases which would assist him in retrieving additional relevant records. A thesaurus may also show hierarchical relationships among words. The user will be able to reference this list to find those words or phrases which are related to the same topic but are more specific or more general in nature. A thesaurus could be constructed and access to it provided for the user to determine the general nature of topics covered in that file and, thus, serve as a "jumping-off-place" for his search.

INDEX REFERENCE

The user will have the capability to list indexes and use the results to formulate more accurate search requests. Also provided will be an item count corresponding to each index term.

TRUNCATION OF SEARCH TERMS

Another facility which will be helpful to the user at the time of formulating his request is the ability to truncate search terms. This facility will enable him to use words without suffixes, thus retrieving records from a file in which various forms of the word are contained. For example, in the request:

FIND TITLE WORK#

the '#' symbol has been used to signify truncation. Assuming the TITLE data element had been indexed for the file being accessed, the records with titles containing the words WORKS, WORKING and WORKED would be retrieved. Truncation also may be used where the spelling of a term is doubtful as:

EMPLOYEE HAN#

Employee records with surnames HANLEY and HANDLEY would be retrieved. The user may then be more specific once he has determined which record satisfies his interest.

A facility similar to truncation will provide for alternative spellings. A search term would be specified with 'don't care' indicators, as in the example below:

EMPLOYEE HANS#N

The ambiguous '#' would cause employee records with surnames HANSEN and HANSON to be retrieved. This would be useful in cases where the exact spelling is unknown. It would be necessary, however to specify at least the first three letters of the name before inserting 'don't care' characters. Truncation options will be provided for searching name, title word, and topic indexes.

SAVE-REUSE FACILITY

A save and re-use facility will be available. At any point within his search request, the user may save the results of his query for later use. He may also save and re-use the request itself.

STANDING REQUESTS

Users may be only interested in any new information which has been added to a file. The standing request facility will be helpful here. Users need only formulate their requests once and leave them with the system. Information which is being added to a file will be passed against the requests and any matching records delivered to the requester.

RECOVERY OF SEARCH RESULTS

If something happens within the system causing interruption of normal service, users should be restored to their place in the search. This should be the responsibility of the system and not the users.

8.12 Output Requests

GENERAL

SPIRES will accept output requests which allow selection within the following options: media, format, document selection, sorting, and generalized report format/content.

OUTPUT MEDIA

The system will provide a spectrum of output media from which a user may choose one or several - appropriate in terms of cost, output volume, convenience (usability), and reusability (machine-readability).

If his output volume is low, the on-line user may be satisfied to accept it from the terminal communication devices: typewriter or CRT. The typewriter supplies him with a hard copy whereas the CRT does not. Since the typewriter is relatively slow and only one line may be listed at a time, flexibility provided via this device will be minimal. The CRT can display several lines at a time, thus providing better formatting and giving the user a scanning facility. The capabilities of the CRT will allow the user to browse through a set of selected records at his own pace.

If his output volume is high or he desires a permanent copy, he can divert it to an off-line batch process: to either a high-speed line printer or computer output microfilmer (COM). The printer output format can be varied in the forms or print chain used, and the number of copies prepared. The microfilm option has three advantages over the printer option: the microfilm requires little storage space, it can be searched and viewed manually or mechanically, and it can be used to produce unlimited hard copy at a small percentage of the cost.

Finally, if his output data must be re-read by the computer at a later time, he can choose magnetic tape, magnetic disk, or punched cards as his output medium. Information stored in this way can also be subsequently listed or distributed externally, e.g. sending a tape to another institution.

OUTPUT FORMATS

Information may be presented in various forms. The user will have a choice in the data elements in each record he wants to see and the sequence in which those elements are to be presented. If he creates a format which he will want to use at another session, he will be able to save the specification and re-use it later.

There are three sources of formats:

1. System-wide standard
2. File standard
3. User-defined

All three sources will be available to the user. He will be notified if he has used a format which is inappropriate for the file in which he is currently working.

The user will be able to set tabs at his typewriter terminal to affect column assignments and margins, set a line length to limit the number of characters to be presented on a line, set a page or screen length for number of lines, and request labels attached to the elements presented. The formatting features provided at the terminal will be limited and straightforward because of the excessive time required to produce sophisticated output on-line.

SELECTING DOCUMENTS FOR OUTPUT

At the time a user asks for the contents of selected records to be listed at his terminal, he will be able to:

1. Specify a range of records or a selection of records to be presented, for example:

   ```
   TYPE  1-5,10,15
   ```

   where only those items indicated would be presented, skipping the rest of the set;

2. Ask for all records in sequence beginning with the first;

3. Ask to be given an option after viewing each record, which permits its storage for later use.

4. Interrupt the listing at any time and,

   a. resume with the interrupted line,
   b. skip to the next record,
   c. skip to a specific record,
   d. skip to the end,
   e. leave the output process entirely,
   f. leave the process temporarily, and return later.

OUTPUT SORTING

Another process concerning presentation is the facility for sorting on one or more data elements. For example, personnel records may be sorted alphabetically by employees' surnames.

DECODING DATA ELEMENTS FOR OUTPUT

If data elements have been stored in coded form, the user has a choice of seeing the information in its compact or expanded form.

REPORT GENERATOR CAPABILITY

A report generator will be provided to allow the user to produce batch listings of selected data base elements in formats of his design.

8.2 FILE MANAGEMENT

8.21 General

There are several needs to be filled in the area of file management. The first of these is a facility for a file manager to define the characteristics of his file without requiring the aid of a programmer. He should be able to enter the specification of the characteristics through a terminal in a non-technical language. Further, the file definition facility must give as much aid as possible in diagnosing errors in the specification. It also must have the capability of allowing the manager to make reasonable alterations to characteristics after the file has been built without having to completely re-build the file.

8.22 Establishment of Files

STORAGE SPECIFICATION

The first characteristic to be specified by the file manager is the amount of direct access storage that will be required for the file. This estimate will be based on the amount of data to be entered in the initial buildup of the file, the rate of growth, and the indexes chosen. The initial allocation of storage should be sufficient to hold the initial data plus the additions which will accumulate over a period of several months. The system must be able to extend the storage for any given file either automatically when the previous allotment has been exhausted or on the entry of a simple command by the manager. If the latter alternative is implemented, the system should issue a warning when the data in the file approaches the current storage capacity.

SPECIFICATION OF DATA ELEMENT ATTRIBUTES

The file manager must decide how to separate the documents into data elements and specify their properties. The properties to be specified are: element name, abbreviations and synonyms, multiplicity, element size, data type, editing functions, and any hierarchical relationship to other elements. The name of an element may contain any of the characters on a terminal keyboard but, for retrieval purposes, an abbreviation must be specified. If it is not, the system will create one. The element size is the number of characters contained in a fixed-length element or the maximum number of characters for a variable-length element. The system must support the following data types: numeric, data, personal name, alphabetic, coded, and full text. Other data types which might be supported are: monetary, linear measures, weights, fractions, and sets of related numbers. Standard sets of data element characteristics may be maintained by the system. Thus, a file manager may elect a default of one or all of these, if it applies to his file.

HIERARCHICAL RELATION BETWEEN DATA ELEMENTS

The concept of a hierarchical relation can best be described with an example. Suppose a file was established with each record being the description of a piece of electronic equipment. Each piece of equipment might be composed of a set of components. One data element might contain the identifications for each of the components. Associated with each value of the component ID element would be an element containing a list of part numbers for that component. Associated with each part there would be an element containing the price of the part. Another example to

2

illustrate hierarchical relations is provided by a bibliographic file where each record represents the reference material for the preprint of a scientific paper. One data element would contain a list of authors of the paper. For each author, one element would contain the organizations with which he is affiliated and an associated element would indicate his mailing address at that institution. Still another element might contain his title with that institution. The system should not place an arbitrary limit on the number of these relationships that may exist among the data elements of any given file.

SPECIFICATION OF INDEXES

Since the manner and degree in which the file is indexed is vital to the retrieval capability that the users will have in accessing that file, the facility given the manager for tailoring the indexing to the requirements of his particular file is extremely important. He should be allowed to specify indexing for any combination of data elements and to have values of more than one element entered into a single index. In addition, it should be possible to add a new index or delete an existing one after the file has been built.

SPECIFICATION OF EDITING RULES

The manager must have the capability to specify editing for the values of an element to be placed in a principal file. Normally, this consists of making selections from a standard set of editing procedures, e.g., function words (like THE, OR, BUT) may be excluded from an index. The manager should also be allowed to specify special editing procedures although he may be required to pay for any programming costs associated with them.

An additional facility would require that the presence (or absence) of a value for one data element necessitates the presence (or absence) of a value for some other element.

DICTIONARY/THESAURUS SPECIFICATION

The File Manager will have the capability to define dictionaries which are specific to a particular set of files. The definition will be part of the file characteristics placed in the system by the File Manager preceding the initial file buildup. A similar capability will exist for Thesauri.

FORMAT SPECIFICATION

Since a user retrieving from the file should not have to specify the format in which information will be displayed on his terminal, some facility is required for assigning standard formats to the file. These formats may be selected from a set provided by the system or the file manager may define some to meet specific requirements of his file.

8.23 File Maintenance

UPDATE

It should be possible to carry out the update function in any of three ways: completely on-line, completely on a batch basis or as a combination of the two. In the on-line mode, update requests would be entered via a terminal, immediately checked for errors and the change in the file executed while the user is still at the terminal. Batch updates could be punched on cards and delivered to a computer operator who would then place them into the batch queue. An intermediate alternative would allow the user to enter the requests from a terminal but allow the system to collect them into a batch and place them in a queue for later execution. In all cases, the system will have a facility to list the updates that were executed for the file.

Three categories of update requests are needed. The first is the addition and deletion of records. The second is the addition, deletion and altering of data elements within records. The third is to be able to copy information from one record to another or from one file to another.

In order for a user to be able to specify, with ease and without fear of ambiguity, which record of the file is to be updated, it is necessary to have a data element which contains a unique value. This data element must be indexed and the system should check each entry in that index to insure that it references only one record in the file. Examples of this kind of data element are; social security numer in a personnel file, Library of Congress card number, and part number in a parts inventory file.

8.24 Output for File Managers

In addition to the output needs for the support of the retrieval function, two special outputs will be required by some file managers. The first of these, to be used to augment the on-line services or to disseminate externally, consists of catalog cards or shelf lists. These must be sortable on at least one data element.

The second output will consist of various statistical descriptions of the file as prescribed by the manager and gathered by the system. These statistics will aid him in predicting the growth of the file, in determining the utility of an index and in various other management tasks. Examples of statistics he might need are; average lengths of data elements, number of times an index is used in single or cumulative retrieval requests, or quantity of each kind of error made in update requests.

8.25 Training

Several facilities will be needed for training file managers and persons who will be assisting them in maintaining files. A consulting service will be necessary for the dual purpose of aiding the managers to establish their files and helping the file maintenance people when they have difficulties with the update function. In addition, classes should be given from time to time to introduce newcomers to the capabilities of the system.

Several kinds of reference material should be written and made available. These are: a primer, a complete file management reference manual, a short version of the reference manual for maintenance people, and reference cards. The last would be very brief excerpts from the manual printed on cards. They would serve principally as reminders to users while on the terminal.

Once a user is communicating with the system, various online aids should be available. He should be able to ask for a brief introduction to the facilities for file management, for examples of the use of these facilities, for explanation of particular terms and prompts, and for an explanation of what responses are available to him.

8.26 Individuation of Retrieval

In the sections above, much attention has been given to facilities to enable a file manager to individualize his file and tailor it to suit his information and retrieval requirements. In addition, it would be useful for the system to provide certain facilities for individualizing the retrieval function to the habits and idiosyncrasies of a particular searcher. The facilities which might be implemented for this purpose are: macros, subset indexes, subset language, unobtrusive observation, and service priority.

SEARCH MACROS

In the macro facility, the user would be able to combine several requests into one and assign a name to it. Subsequently, he could cause the set of requests to execute by entering the macro name. This feature would reduce the effort of users who repeatedly carry out some particular sequence of requests. For example, suppose someone frequently entered some term of a topic index, requested all synonyms, assembled these into a retrieval request for all records containing any one of them and finally requested to look at the first three of the retrieved records. If this were all combined into a macro, it would save him a significant amount of keying and possibly some mistakes.

SUBSET INDEXES

At times, some user might wish to do exhaustive searching though part of a very large file. For example, a geologist might wish to work with the section of an earth sciences bibliographic file which pertains to precious metals. In order to reduce the cost of the on-line retrieval, it would be advantageous for him to be able to request the creation of a file which would be a subset of the full earth sciences file. To achieve minimal cost, the file records themselves would not be duplicated, but rather, a separate set of smaller indexes would be built.

LANGUAGE SUBSETS

In order to make the process of entering retrieval requests simpler and thus reduce both the amount of learning required and the number of errors made, the system should support language subsets. A user would only need to learn those request formats which apply to his individual needs.

UNOBTRUSIVE OBSERVATION OF USER HABITS

Most users will probably make certain errors quite frequently. If a record were maintained by the system of each user's habits, then, for those errors which are made consistently (and also corrected each time by the user), the system could make the correction for the user. This facility should, however, be an optional one.

USER PRIORITY

Normally the system will consider the requests of all users to be of equal importance and will optimize the servicing of requests to keep the average response time to a minimum. However, on some occasions, a particular user may have need for faster service and be willing to pay for it. Thus the system should provide the facility for a user to assign priority for his requests and to charge him higher rates accordingly.

9.0 Generalized Search and Retrieval First Implementation Scope

In order to fully understand this section, it is necessary to have read Chapter 8.

The system will have the following general characteristics:

1. Flexibility - the system must be able to accommodate a variety of files, including any of the bibliographic data available in machine-readable form.
2. Adaptability - it must be possible for a user to use and be charged for only that part of the system which he needs.
3. Modifiability - the system should be designed and implemented in such a way that it is easy to change. In particular, it is foreseen that the interactive search may require expansion.

9.1 Retrieval

The following search facilities will be implemented:

1. indexes - the user will be able to use indexes of the following types in his search requests. However, for any given file, he may use only the indexes associated with that file.
    a. personal name
    b. title word
    c. topic - contains terms descriptive of the subject matter of documents in a bibliographic file.
    d. numerical
    e. date
    f. coded
    g. file partition - ability to divide a file into sections. For instance, a file of physics papers might be partitioned into experimental, theoretical and survey sections.
    h. user-defined - the data type, editing or format is specified by a file manager especially for his file. If any additional implementation cost is

required, it will be at his expense.
   i. citation
2. access via non-indexed elements
3. on-line search
4. batch search
5. query language
   a. logical expression - several simple requests may be combined into one request by use of the words: AND, OR, NOT. For example: FIND AUTHOR Smith AND TITLE Hemophilia.

   b. weighted terms- each term of a request may be assigned a number by the user. Only those records which score the same or more than he specifies will be retrieved.
   c. interactive
6. dictionaries
   a. exclusion - contains list of terms which will not be put into an index.
   b. synonym
7. index reference - the ability to inquire as to what values are in a particular index
8. save and re-use - the ability to name a search request or the results of a search and have the system store it. The request or results could be used later upon entry of the name assigned.
9. standing request - the ability to enter a retrieval request and have all new material added to a file compared with it. Any records meeting its criteria would be communicated to the user.
10. on-line recovery of search process - to insure that a user will not lose the results of an interactive search derived over a set of several interactions because of a temporary system failure.

The output facilities to be implemented are:

1. on-line
2. batch print
3. batch tape
4. formats
   a. system standard - formats specified by the system and available for anyone's use.
   b. file standard - formats specified by a file manager and available to any user of that file.
   c. user-defined - the ability for a user to specify a format while at the terminal.

5. sorting (for batch output only) - the ability to list retrieved records on a printer, ordered on the values of one or more data elements.
6. catalog cards - a printing, directly onto cards, of information contained in selected elements of a bibliographic file. This would be a batch operation.

The following training facilities will be provided:

1. reference manuals
2. reference cards
3. on-line aids - capability for the user to ask for help from the system through his terminal.

9.2 File Management

The following major facilities will be implemented:

1. definition of file characteristics
2. modification of file characteristics
3. buildup of file from initial data
4. updating
5. special listings - these will generally be unique to a file as specified by the file manager
6. statistical feedback
7. training

The file definition facility will allow the file manager to specify:

1. amount of required storage - ability to specify to the system the initial size of the file and its rate of growth.
2. data elements
   a. element name
   b. multiplicity
   c. element size
   d. data type - e.g., dates, personal names, numbers.
   e. choice of input editing
   f. hierarchical relations - for instance, one data element might contain a list of project names. Associated with each project is a data element which has a list of employees assigned to it. Associated with each employee is a data element which contains a list of tasks for him.
   g. automatic functions (to be executed upon occurrence of transaction for the element)

3. indexing
   a. which elements will be indexed
   b. addition, deletion of indexes
   c. editing of values to be indexed
4. dictionaries
   a. codes
   b. exclusion - the user should be able to override and force a term into the index for some records.
   c. inclusion - a list of words which will be put into an index. All other words will be omitted from the index.
   d. synonym dictionaries
5. display formats - ability to specify standard formats for the file. Each format would have a name which a user would enter in an output request. The format would specify the elements to be displayed and their order.
6. error severity level - the ability for the manager to specify the action to be taken upon the occurrence of various errors. The choice of actions includes: nullifying the user's request, presenting an error message and attempting to correct the error.

The following file maintenance facilities will be provided:

1. tape conversions
2. on-line entry of input and update requests
3. batch execution of updates
4. on-line execution of updates
5. update requests
   a. addition, deletion of records
   b. addition, deletion, alteration of elements or parts of elements.
   c. copy - from record to record and from file to file.
6. index of the record identification data element
7. applications - specific batch facilities will be provided on demand when feasible. These will normally be paid for by the user who requests them.
8. File merging and elimination of duplicates.

The training facilities which will be provided are:

1. reference manuals
2. reference cards
3. on-line help
4. consultation service

Miscellaneous Features

1. file specific message of the day
2. collection facility for user documents submitted on-line

PART IV

SHARED FACILITIES

10.0 SUMMARY OF CURRENT SHARED FACILITIES

10.1 General Concepts

DEFINITION

Shared facilities consist of software and hardware designed to provide concurrent service to functionally related applications.

ECONOMIC CONSIDERATIONS

A gross estimate reveals that in terms of implementation effort, SPIRES/BALLOTS II may be broken down approximately as follows:

... BALLOTS - 1/3
... SPIRES - 1/3
... Shared facilities - 1/3

If each application user pays for his own development plus half for the shared facilities, that user effectively gets the use of sixty-seven percent of the system for half the total investment. Alternatively, if two users invest similar amounts in separate development efforts, each is given substantially less for his money. Another operative factor is hardware economy of scale. If two users pool their resources to acquire shared hardware, the resulting individual capability will be greater than it would with separate installations. This simple analysis argues for continuing combined SPIRES/BALLOTS development.

10.2 Present Shared Facilities

COMPUTER OPERATIONS ENVIRONMENT

SPIRES/BALLOTS I software executes on an IBM 360 Model 67 located in the Campus Facility of the Stanford Computation Center. This computer has one million characters of main storage, and processes data input and output through ultra-high-speed and high-speed direct-access devices as well as magnetic tapes, card equipment, and line printers.

Installation software and procedures are directed toward a rapid throughput computation-oriented market. Although the data processing facilities provided are of excellent quality, high priority is placed on keeping the computation facilities operative. If a file failure occurs, correction must wait until a scheduled software maintenance interval. This could result in an unacceptable inconvenience to the non-standard user who has very large, continually updated files.

There are two pieces of computer memory available for program execution. The first is approximately 100,000 characters long, and will accept no job whose duration exceeds two minutes. The second is approximately 300,000 characters long, and will accept jobs of any duration. SPIRES/BALLOTS I uses the latter. A great disadvantage is that while someone else is executing in this portion of memory, SPIRES/BALLOTS cannot and vice versa. This precludes extended, exclusive use of the computer resources by SPIRES/BALLOTS I.

The policy in this operations environment is to discourage long-duration jobs by charging them more per execution minute as the job progresses in time on the computer. A further discrimination is made between day and night jobs; it is cheaper to run at night. It is clear that these policies are not constructed to benefit a system such as SPIRES/BALLOTS I. A further problem is a lack of guaranteed access to the system from a terminal; there are over 200 terminals connected to the system and only 60 can be in use simultaneously.

The model 67 is currently approaching its capacity, at least during peak periods. These periods occur near mid-term and final examination time or roughly eight times per year. During such intervals the execution backlog grows long, and it is difficult to gain access to the system through a terminal.

ON-LINE EXECUTIVE PROGRAM

The SPIRES/BALLOTS I Supervisor is an on-line executive program designed and developed by project personnel to service several on-line users simultaneously. The purpose of an on-line executive program is to regulate the competition for service and resources among several terminal users. The program attempts to insure that each user gets a reasonable share of available execution time. Experience with the SPIRES/BALLOTS I supervisor has demonstrated the feasibility of the approach taken; response time averages three seconds for simple search requests.

TERMINAL HANDLER

The terminal handler performs the actual input/output operations between remote terminal locations and the main computer. Its role is that of a middleman standing between the terminal lines and the on-line executive program. This function is currently discharged by MILTEN, a program provided by the Campus Facility installation. Part of the program resides in the main storage of the Model 67, and the rest in a smaller computer (PDP-9) to which the terminal lines are attached.

ON-LINE DATA COLLECTOR/TEXT EDITOR

The purpose of this program is to allow the on-line collection of input data for later use by batch computer runs. It further allows correction and modification of such data at the character level. This facility has been found extremely useful in gathering data to be used in file building; most users have chosen it in lieu of punched cards and found it easier and cheaper than less flexible alternatives.

The need for a Data Collector/Text Editor is currently satisfied by WYLBUR, which is part of the Campus Facility installation software. It has been found to be excellent in all respects save one: it requires the user to backup his files, rather than provide such service automatically.

FILE SUPPORT

The basis for any information storage and retrieval system is the collection of files it handles. These files may or may not have any connection among themselves. For example, the entire collection may contain files related to personnel records, medical data, or bibliographic data concerning published documents. There is no restriction on the information that can be stored and no two distinct groups of files need have a relationship.

Files within the collection that are connected or related to one another in some predetermined way are defined to be a set of related files. The system supports two types of related files: principal and statistical.

Principal files serve as the basis of operation for the user within the system. In these he accumulates his primary data: texts, abstracts or other data elements, their associated access indexes, and file characteristics.

Statistical files contain information on the contents and usage of corresponding principal files.

RECOVERY/RELIABILITY

The Campus Facility System fails at least once every 36 hours, and sometimes more often. The incidence of failure may seem high, but realistically speaking, the system has excellent reliability for such a complex collection of facilities. Such failures, however, can cause an unacceptable loss of a large continually updated file.

Recovery of files whose integrity has been lost in such situations is accomplished by periodically copying the file to magnetic tape (called dumping) and recopying back to disk (called restoring) following the failure. It has proved economical to dump a file after each one-hour aggregate of file building time.

AVAILABILITY

The current SPIRES/BALLOTS files are available during the day and most of the night. The on-line executive program, however, is not. At the present time, there is no regularly scheduled SPIRES/BALLOTS service block, and users must bring SPIRES/BALLOTS into execution themselves. As discussed above, they pay premium prices as a result.

11.0 LONG-RANGE SCOPE, SHARED FACILITIES

BALLOTS and SPIRES will share common software/hardware facilities. It is difficult to predict the nature of application areas to be added in the future. In theory, any new application requiring on-line storage and manipulation of data can be accomodated. A necessity therefore exists to implement all shared facilities in a generalized, modular fashion to facilitate additions at the application level.

With the exception of added utility programs, there will be little expansion of shared facilities beyond the SPIRES/ BALLOTS II effort. Applications added later will be designed to interface with SPIRES/BALLOTS shared facilities, and will cause few perturbations at the shared facility level.

It follows that the long-range scope is identical to the scope for implementation in 1970-71.

12.0 FIRST IMPLEMENTATION SCOPE, SHARED FACILITIES

Below is a list of those facilities whose sharability is certain. As the detailed analysis and general design phases proceed, it may become apparent that other facilities may be generalized and shared (e.g., a batch update that works for both library and GISR users). Since no certainty now exists with regard to such facilities, they are treated separately in the two preceding sections.

COMPUTER OPERATIONS ENVIRONMENT

The operations environment for SPIRES/BALLOTS II will be a Data Facility. The hardware chosen will be only large enough to service present applications, with later augmentation as growth dictates. Procedural orientation within the facility will emphasize data handling rather than computation. High priority will be placed on the recovery of lost data as well as resumption of service to other users.

The Data Facility will handle long-duration and non-terminating jobs as well as short-duration utility jobs. There will be a greater guarantee of access to the machine during normal working hours, and machine resources will be provided once access is gained. Since the pressure of dominant, cyclic workloads will be absent, access contention will exist only within the data facility user group.

ON-LINE EXECUTIVE PROGRAM

All services provided by the SPIRES/BALLOTS I Supervisor will be provided in SPIRES/BALLOTS II. Design goals will include a maximum of flexibility and generality to facilitate the addition of new applications. Another desired feature is changeability of the user command language without resort to reprogramming. The language must be augmentable through the addition of new applications as well as changeable to whatever new experience dictates.

TERMINAL HANDLER

All services now provided by MILTEN running in the Model 67 and the PDP-9 will be provided by the new system. This could happen through the adaptation of MILTEN or some other pre-existing package to the new environment.

One additional condition to be met is the accessability of the data facility not only through new data facility terminals (CRT's, CRT's with hard copy, and 2741 typewriters) but also through the present campus communications network (2741's presently installed and hooked to the Campus Facility).

ON-LINE DATA COLLECTOR/TEXT EDITOR

All facilities now provided by WILBUR will exist as part of the new shared facilities. As with the terminal handler, this could happen through the adaptation of Campus Facility software, IBM software, or some presently unknown alternative. An additional feature will be the use of the text-editing capability in conjunction with on-line updating of data files.

FILE SUPPORT

The system will support, in addition to the principal and statistical files mentioned in 10.0, two other file categories: historical and holding.

Historical files are of two types. The first includes accumulations of transaction records that have updated principal files. Their role in file recovery is described below. The second type captures records deleted from principal files. This provides an alternative to the re-keyboarding of deleted records when their reuse becomes desirable. Both types of files will generally be retained as magnetic tape files.

Holding files are temporary files of data selected from principal files. These will fulfill the input requirements of scheduled batch processes or satisfy individual standing requests from users for selective reporting.

RECOVERY/RELIABILITY

Since files are the basis of the system, their reliability is extremely important. Information should not be irrecoverably lost or damaged in any way by user error, machine malfunction, or program problems. Should a file become damaged or destroyed, a set of methods must exist for immediately re-creating an image of the file as it was just prior to the malfunction, and quickly restoring service. The following discussion describes two techniques that will be used to achieve this.

1. SIMPLE COPY/RESTORE At specified intervals, a set of files is copied to magnetic tape. If the on-line version of those files suffers damage or is lost, the magnetic tapes can be recopied back on-line, thus restoring the files to their status as of the last copy to tape. In cases where few updates have occured in the intervening period, this method may be sufficient providing absolute file integrity is not required.

2. COPY/RESTART This method is similar to the simple copy/restore, with one enhancement: the history file, containing all adds, deletes, and changes to the file since the last copy, will be used to update the restored version to the condition of the file just prior to the malfunction. This is done when a file has undergone many changes since being copied to tape, and absolute file integrity is required.

AVAILABILITY/SECURITY

The availability of file sets has several aspects: service hours, public vs. private files, multiple users of files, and file security. All file sets and all information within those sets are not available to everyone at all times. Some files may be available for on-line retrieval at specified times during a day (if those files are on-line during that time) and perhaps available for batch maintenance at some other time. Other files may be concurrently available for retrieval and maintenance, implying on-line maintenance. There may be another category of files which are kept off-line and only placed on-line at the request of the user.

The availability of files to the user community also depends upon the status (public or private) which has been defined for those files. Public files can be accessed by anyone who desires to obtain information from them. Some large public files contain information received from a national bibliographic service via magnetic tape . A file may belong to a particular user who maintains the file and has complete responsibility for it. Such a file may be termed a personal file and still be available publicly, e.g. bibliographic data regarding a professor's private library. Private files can be accessed by a restricted number of users, possibly only the person responsible for that file. There are several variations on the public/private concept. Access to a file may be unrestricted; changing data within the file may be restricted to one or a few persons and still allow unrestricted query. Alternatively, access to a file may be partially restricted such that only a portion of a file or a certain set of data elements is available to general users.

There may be several users of the entire system at any one time. If a file is available to more than one user, there may be two or more users accessing information from the same file simultaneously. One user is not refused access to information in a file because information in that file is already being accessed by another (unless both users are attempting to update at the same time).

The ability to maintain files as public, private, or semiprivate is dependent upon a file security facility. Security must exist at these levels:

1. Files must be secured against access by anyone not having authorization.

2. Specified data elements within a file must be secured against access by anyone not having authorization.

3. Files must be secured against modification by anyone other than the file manager or persons given authorization by him.

Security at all levels could be effected through the use of group or individual passwords. A password is a string of characters which has been specified by the file manager as a key to gain access to his file. A searcher not responding to a request for the correct password would be denied his request for information retrieval. Other implementation possibilities include user definition of a security algorithm appropriate to a particular set of files.

ACCOUNTING

It will be necessary to design and implement accounting software to gather information for customer billing. This software must be sophisticated enough to distinguish between a user whose support requirements are small, and one who has complex requirements. Customer charges must accurately reflect machine resources actually utilized. With the exception of overhead rates, there will be no hidden subsidy of expensive facilities by customers not actually using them.

Such software is difficult to implement. This fact is reflected by a general lack of vendor accounting support until recently. In spite of this fact, it may be possible to adapt software developed elsewhere for this purpose, such as the System Management Facilities package distributed by IBM.

CHARACTER SETS AND SYMBOL REPRESENTATION

The capability will be provided to display or transliterate special symbols; for example:

... Mathematical symbols
... Symbols used in the physical sciences
... Greek letters
... Diacritical marks

Wherever direct display is not feasible, a notation such as 'A = *ALPHA*', could be used.

REPORT GENERATION

The capability (consistent with security) to select, format, and list data base elements will be provided on a batch basis.

APPENDIX A

GLOSSARY

Definitions of library terms in this glossary have been made consistent with those in the "Anglo-American Cataloging Rules" and the "A.L.A. Glossary of Library Terms" whenever possible.

Any word which is followed by an asterisk (*) in a definition is itself defined in this glossary.

AACR--Anglo-American Cataloging Rules. A standard reference book of rules used in cataloging.

ACCESS POINT--An entry route into a file*. The only access point into a sequential file* is the beginning of the file. An access point into a direct access file* may lead directly to the desired record*. In order to facilitate searching* indexes* are constructed to gather together access points to records with a common data element value*.

ACCESSION--(n.) A book or other similar material acquired by a library for its collections. (v.) To record, in the order of acquisition, books and other similar material added to a library's collections.

ACQUISITION--The acquiring of books, periodicals*, and other material by purchase, exchange*, and gift, and the maintenance of necessary records of these additions.

ACQUISITION DEPARTMENT--The administrative unit in charge of acquiring books, periodicals* and other material by purchase, exchange*, and gift and of keeping the necessary records of these additions. In the Stanford University Libraries the Acquisition Department includes the Order Division, Serial Records Division, Binding and Finishing Division*, Gift Division, and Exchange Division.

ADDED COPY--At Stanford, a duplicate of material already in the SUL* System, if it is added, or to be added to the system.

ADDED ENTRY--An entry*, in addition to the main entry*, under which a bibliographical entity is represented in a catalog*; a secondary entry, including subject entries*.

ADP--Administrative Data Processing. A computer facility which is a part of the Stanford University Controller's Office. ADP currently has an IBM System 360 model 40 computer.

ALPHANUMERIC DATA--Data which may be made up of letters, numbers, punctuation marks, or any combination of the preceding.

ANALYTIC--See analytical entry*.

ANALYTICAL ENTRY--An entry* for a work or part of a work that is contained within a collection, series*, or other bibliographical unit for which another, comprehensive entry has been made.

ANOTHER EDITION--An edition of a work acquired by a library that differs from other editions of the same work already in the library.

ARCHIVES--1. An organized body of documents or records relating to the activities, rights, claims, treaties, constitutions, etc., of a family, corporation, community, nation, or historical figure. 2. A place where such records or documents are kept.

ARREARAGES--Specifically used to refer to the backlog of books not yet cataloged.

AUTHOR--The person or corporate body chiefly responsible for the creation of the intellectual or artistic content of a work, e.g., the writer of a book, the compiler of a bibliography, the composer of a musical work, the artist who paints a picture, the photographer who takes a photograph.

AUTHOR ENTRY--The entry* of a work in a catalog* under its author's* name as heading*, whether this be a main or an added entry. The author entry may consist of a personal or a corporate name or some substitute for it, e.g., initials, pseudonym.

AUTHOR-TITLE CATALOG--A catalog* consisting of author and title entries, and sometimes entries* for editors, translators, series*, etc., but excluding subject entries*.

AUTHORITY LIST or FILE--An official list of forms used as headings* in a catalog*, giving for author* and corporate names, and for the forms of entry* of anonymous classics the sources used for establishing* the forms, together with a record of cross-references and/or history cards made; an official list of topical subject* headings used in a catalog and a record of cross-references made.

AUTONOMOUS LIBRARY--See Coordinate Library*.

BACKGROUND PROCESSING--Computer processing which takes place when the on-line system* has no requests to service.

BACKUP FILES--Copies of files which are maintained for use in the event of damage to the original file*.

BALLOTS--Bibliographic Automation of Large Libraries on a Time-Sharing System. Acronymn for the Library Automation project.

BATCH RETRIEVAL--Requests are accumulated by a computer operator or by the system and placed in a queue* to be run as a group. The output* is listed on a printer* and returned to the user some time after he makes the request.

BIBLIOGRAPHIC FILE--A file* consisting of records* containing data elements* such as author, title, date published, number of pages, catalog number.

BNB--British National Bibliography. See National Bibliography*.

BINDING--1. The process of producing a single volume* from leaves, sheets, signatures, or issues of periodicals*, or of covering such a volume*. 2. The finished work produced by this process. 3. The cover of a volume.

BINDING AND FINISHING DIVISION--The division of the Acquisition Department* responsible for labeling, plating, pasting in of pockets, binding preparation, binding, and repair* of books, periodicals*, pamphlets, etc.

BLANKET ORDER--An order placed with a dealer to ship material in specified subject areas with the understanding that all such material will be accepted by the Order Division unless it is a duplicate of material already in the collection.

BOOK CATALOG--A catalog* in book form rather than in card form.

BOOK NUMBER--A designation, consisting of letters and numbers, which uniquely identifies a work among other works with the same classification number*. Usually the second of a call number*, coming after the classification number*.

CALL NUMBER--Letters, figures, and symbols, separate or in combination, assigned to a book to indicate its location on the library shelves. It usually consists of a classification number* and book number*.

CARD CATALOG--A catalog* in which entries* are on separate cards arranged in a definite order in drawers.

CARD NUMBER--A number, or combination of a letter, letters, or a date, and a number, that identifies a particular card in a stock of printed catalog cards*.

CATALOG--A list of books, maps, etc., arranged according to some definite plan. As distinguished from a bibliography, it is a list which records, describes, and indexes the resources of a collection, a library, or a group of libraries. In practice, some catalogs also contain records for items which are on order and items which are in the cataloging* process.

CATALOG CARD--1. One of the cards composing a card catalog*. 2. A plain or a ruled card, generally of standard size, 7.5 cm. high and 12.5 cm. wide, to be used for recording entries* in a catalog*.

CATALOGING--The process of preparing a catalog*, or entries* for a catalog. In a broad sense, all the processes connected with the preparation and maintenance of a catalog, including the classification of books and the assignment of subject headings*. In a narrower sense, the determining of the forms of entry* and preparing the bibliographical descriptions for a catalog*.

CATALOG DEPARTMENT--1. The administrative unit of a library in charge of classifying books and preparing the catalog. 2. The library quarters where the cataloging processes take place.

CHARGE--A record of the removal of a book from the library stack*, usually as a loan to a patron, less often for internal library processing.

CHARGE FILE--1. A record of books loaned, usually consisting of records arranged by date or call number*. Also called a circulation* file. 2. The physical file.

CHECK-IN--See Serial Check-In

CIRCULATION--1. The activity of a library in lending books to borrowers and keeping records of the loans. 2. The total number of volumes*, including pamphlets and periodicals* loaned during a given period.

CIRCULATION DESK--A counter or desk where books are loaned and returned, and where records of this activity are kept.

CLASSIFICATION NUMBER--1. A number, or combination of numbers and letters, used to designate a specific element of a classification scheme. 2. The notation added to a book and to its entry* in a catalog* to show the class to which it belongs. The first element of a call number*.

CLASSIFICATION SCHEDULE--The printed scheme of a particular system of classification, such as a Library of Congress Classification Schedule.

CLEARINGHOUSE--A center set up to collect and disseminate information pertaining to some discipline.

CLOSED ENTRY--An entry* with completed bibliographical information covering all parts of a multivolume work, viz., a complete set*.

COMPOUND SEARCH REQUEST--A set of simple search requests* connected by words such as AND, OR, or NOT.

CONTINUATION FILE--A list of serials*, sets* appearing at irregular intervals, and books in series*, recording numbers and parts received.

CONVERSION--The translation of data, under computer control, from one format* to another.

COORDINATE LIBRARY--A library on the Stanford University campus which operates independently and does not come under the administration of the Stanford University Libraries. Specifically, the Food Research Institute Library, Hoover Institution Library, Jackson Library of Business, Lane Medical Library, Law Library, and Stanford Linear Accelerator Center Library. See Appendix D for a complete list of the libraries at Stanford.

CRITERIA--The conditions, stated by a user in a search request*, which data in a record must meet to be retrieved.

CRT--Cathode Ray Tube--A computer terminal* which is like the visual part of a television set with a keyboard added.

CSt--An internationally recognized symbol for Stanford University (California - Stanford) used in such publications as the National Union Catalog* and the Union List of Serials*.

CST--Cutter-Sanborn Table. A three figure alphabetical order scheme, an alteration of the two-figure Cutter table*, made by Kate E. Sanborn.

CUTTER TABLE--Either the two or the three figure alphabetical order scheme developed by C.A. Cutter which provides decimal numbers that can be combined with the first letter of surnames or other words to order and uniquely identify books under a given classification number*. Also referred to as Author Table.

DATA BASE--See file*.

DATA ELEMENT--A part of a record*. For instance, in a personnel file, a record may be made up of data elements containing the employee's name, age, position, salary, and date of employment.

DATA ELEMENT VALUE--The information stored in a data element*. For instance, the value of the data element "author" might be "Jones" in record* 92 but "Smedley" in record 567.

DATA SET--A file stored on a disk pack* and accessable by WYLBUR*.

DATA TYPE--The nature of the information to be stored in a data element*. For instance, the data type of the data element salary is numeric. Other data types are names of people, dates, and codes.

DELINQUENCY--In circulation: a record keeping designation for bills that have not been paid or material not returned by the end of the academic quarter.

DELINQUENT BILLS--In circulation: a bill sent at the end of a quarter informing the user that his registration will be held by the Registrar until he clears his record with the library.

DESY FILE--A machine-readable* high energy physics file produced and distributed by Deutsches Elektronen Synchrotron.

DEWEY DECIMAL CLASSIFICATION--1. The classification scheme for materials devised by Melvil Dewey, which divides human knowledge into ten main classes, using a notation of numbers, with further decimal subdivisions.

DICTIONARY CATALOG--A catalog* in which all the entries* (author*, title*, subject*, series*, etc.) and their related references are arranged together in one general alphabet. The subarrangement frequently varies from the strictly alphabetical.

DIRECT ACCESS FILE--A file* in which any records* may be retrieved without having to pass over all preceding records.

DISCHARGING--Cancelling the loan record (charge*) for a book when the book is returned to the library.

DISK PACK--A collection of five magnetic disks* that are connected together and treated as one unit.

DISPLAY--Information presented on a CRT terminal*.

DLC--An internationally recognized symbol for the Library of Congress (District of Columbia - Library of Congress) used in such publications as the National Union Catalog* and the Union List of Serials*.

ENTRY--1. A record of a bibliographical entity in a catalog* or list. 2. A heading* under which a record of a bibliographical entity is represented in a catalog or list. See also Heading*.

ERIC FILE--A machine-readable* file containing information about research in educational methods and technology. It is produced and distributed by the Educational Resources Information Centers.

ESTABLISH--The process of determining and verifying the exact and correct form of a catalog entry*.

EXCHANGE--1. The arrangement by which a library sends to another library, institution, or society its own publications or those of an institution with which it is connected and receives in return publications of the other institution, or sends duplicate material from its collection to another library and receives other material in return. 2. A publication given or received through this arrangement.

EXCLUSION LIST--A list containing words which are not placed in some index* when they occur as values of indexed* data elements*. A typical exclusion list might include "a", "an", "and", "the".

FASCICLE--One of the temporary divisions of a work which, for convenience in printing or publication, is issued in small installments, usually incomplete in themselves, which do not necessarily coincide with the formal division into parts*, etc.

FILE--A collection of information, existing on some storage medium and organized in a way that allows segments to be located and extracted in a systematic manner.

FILE CHARACTERISTICS--The properties of a file* which distinguish it from other files in the system. Examples are: the list of data elements*, description of the indexing and output* formats*.

FILE DEFINITION--The process of specifying the file characteristics* for a particular file*.

FILE MAINTENANCE--The process of defining a file*, inserting the first set of records*, updating* the file, and restoring* the file when damaged.

FILE MANAGER--The person responsible for the definition and maintenance of a particular file*.

FILE SEQUENCE(or File Order)--The sequence of records in a file* which is determined by some data element* in each record*. For example, a purchase order file might be sequenced by purchase order number(the data element) and a vendor file might be sequenced by the vendor number (the data element) appearing in each record.

FORMAT--A description of the physical arrangement of information used by programs for entering or putting out data.

FULL CATALOGING--Cataloging* that gives detailed bibliographical information in addition to the description essential for identifying books and locating them in a library.

FULL-TIME EQUIVALENT (F.T.E.)--Any number of people whose hours of working time, when added together, equal one full-time position*.

FULL-TIME POSITION--Normally, a 40-hour work week for each person. When an employee has week-end or night duty, 38 1/2 hours is a work week.

GENERALIZED INFORMATION STORAGE AND RETRIEVAL (GISR)--An information storage and retrieval system to be developed by the SPIRES/BALLOTS project to service the varied needs of the Stanford community.

HARD COPY--A printed copy of machine output in a visually readable form, for example, printed reports, listings, documents, summaries, etc.

HARDWARE--The physical machinery that makes up a computer system.

HEADING--1. A name, word, or phrase placed at the head of a catalog* record to provide a point of access in the catalog. Headings function as entries in the cataloging* of particular bibliographical entities. 2. Sometimes used in descriptive cataloging to denote the aspect concerned with uniform modes of representing the names of persons and corporate bodies and the titles of works in headings. See also Entry*.

HISTORICAL FILE--A file* containing information deleted from a principal file*, or a file which contains the history of recent transactions* affecting a principal file.

HOLDS--(N.) The record keeping designation applied to the process of noting that a user wishes to reserve the next use of some material currently circulating.

HOLDINGS--The books, periodicals, and other material in the possession of a library; a record which is kept, usually of an indefinitely continuing publication, i.e. holdings record; a computer listing of the copies, volumes*, parts*, etc. of a bibliographic items held by a library.

IMPRINT--1. The place, publisher and/or printer of a book (if known), and date of publication of a book. 2. The statement giving such information in a bibliographical description of a printed work. 3. A book or other publication that has been printed.

IMPRINT DATE--The year of publication or printing.

INDEX--An appendix to a file* which contains a list of the values for one of the data elements*. This list is ordered in some manner and enables a user to directly access records* in the file which have a particular value for the data element. One file may have more than one index.

INFORMATION QUESTION--Any question asked of a reference librarian that can be answered immediately without consulting a reference tool. This definition, of course, depends upon the question asked and the reference librarian to whom it is asked. As opposed to reference questions*.

INPUT--The transmission of information from a terminal* or some other device to the computer.

INTERACTIVE RETRIEVAL--The user enters his search request* into the computer from a terminal* and receives a prompt* indicating that query* has been processed. Response time* generally is less than a minute. The user then revises the retrieval request, enters a new request, or requests that the data located by the search be output*.

INTERLIBRARY LOAN--1. A cooperative arrangement among libraries by which one library may borrow material from another library. 2. The loan of library material by one library to another library.

JOB--A specified group of tasks prescribed as a unit of work for a computer.

JOINT AUTHOR--A person who collaborates with one or more associates to produce a work in which the contribution of each is not separable from that of the others.

KEYWORD INDEX--An index* constructed for a data element* that contains values descriptive of the contents of a document. For instance, a keyword index might be constructed for the data element "subject" of a bibliographic file.

LC-0--The mark on an NPAC* notice indicating that no Library of Congress catalog copy was found in pre-order search for material in scope of NPAC.

LC-0-X--The mark on an NPAC* notice indicating that no Library of Congress catalog copy was found for book-in-hand search for material in scope of NPAC*.

LETTERING--1. In binding, the process of marking a cover or spine with title or title or other distinguishing characters, and, in a loose sense, with the accompanying ornamentation. 2. The result of this process.

LIBRARY OF CONGRESS CLASSIFICATION--A system of classification for books developed by the Library of Congress for its collections. It has a notation of letters and figures that allows for alphabetic and/or decimal expansion.

LOG ON or OFF--To become connected to or disconnected from the computer by means of a standard procedure. Logging onto the computer might entail providing one's name and password*.

LOGICAL EXPRESSION--See compound search*.

LOST BOOK--The record keeping designation for books physically lost and for material not returned 60 days after overdue notices are sent.

MACHINE-READABLE--Data which can be read by a device attached to a computer, e.g., punched cards or magnetic tape*.

MAGNETIC DISK--A metal plate on which data can be recorded in tracks analogous to those on a phonographic record. The recording is done magnetically rather than by impression in the surface. The tracks have addresses and may be directly accessed.

MAGNETIC TAPE--Tape similar to the tape used in ordinary tape recorders. Recording on and retrieval from magnetic tape is sequential*.

MAIN ENTRY--1. The prinicipal record usually the author entry*, of a bibliographical entity, presented in the form by which the entity is to be uniformly identified and cited. The main entry normally includes the tracing* of all other headings* under which the record is to be represented in the catalog. 2. The heading under which such a record is represented in the catalog*.

MARC--MAchine-Readable Cataloging. A service of the Library of Congress providing catalog data in machine readable form.

This information is available on magnetic tape to subscribing libraries.

MEMORIAL FUND--A fund of money created by donations for the purchase of books in memory of specified individuals.

MODE--A state of operation. For instance, two computer modes of operation are batch* mode and on-line* mode.

MONOGRAPH--1. A systematic and complete treatise on a particular subject, usually detailed in treatment but not extensive in scope. It need not be bibliographically independant. 2. A work, collection or other writing that is not a serial*.

MONOGRAPHIC SERIES--See Series*.

MOUNT--To connect a disk pack* or magnetic tape* to the computer system.

MULTIPLE ACCESS FILE--A manual or automated file* in which a particular record* may be located if one of several data elements* is known. This is accomplished by creating several indexes* to the file. For instance, a multiple access vendor file might typically enable a user to search by vendor name, vendor number, vendor purchase order number date of purchase or similar data elements.

NATIONAL BIBLIOGRAPHY--A list of works published in a country; or, in an extended sense, of works about a country, by natives of a country, living in that country or elsewhere, or written in the language of a country.

NATIONAL UNION CATALOG--Several sets of reference volumes containing catalog entries for titles cataloged by the Library of Congress or by other cooperating libraries which have agreed to submit catalog entries to the catalog.

NPAC--THE NATIONAL PROGRAM FOR ACQUISITIONS AND CATALOGING. The program in which participating libraries notify the Library of Congress that they are ordering or have received a book for which Library of Congress cataloging information is not currently available. The Library of Congress will then notify the participating library of its intent to catalog or not to catalog this item. The program is limited to acquisitions from certain countries of publication with certain imprint dates. Libraries participating in the NPAC program receive a full depository set of all cards produced by the LC Card Division. Also called Title II-C and Shared Cataloging.

NSA FILE--A machine-readable* high energy physics file produced and distributed by Nuclear Science Abstracts.

NST--NEW SERIAL TITLES. A serial* publication with monthly and quarterly issues and a cumulative annual volume which provides information about new serial titles added to American libraries.

NUC--See National Union Catalog*.

NYPL--New York Public Library.

OFF-LINE--Not in direct contact with the computer.

ON-APPROVAL--A program whereby material is received from vendors on an "on-approval" basis. If, after review, the material is not selected for purchase, it is returned to the vendor.

ON-LINE--In direct contact with the computer.

ON-LINE SYSTEM--A system able to service more than one user simultaneously. The users generally communicate with the system through terminals which are located some distance from the computer.

OPEN ENTRY--A catalog entry* which provides for the addition of information concerning a work whih is still in the process of being published, or about which complete information is lacking.

OPEN SET--An incomplete set* for which the library expects to receive an indefinite number of volumes. As opposed to a Terminal Set*.

ORDER DIVISION--The administrative unit that has charge of acquiring books and other material by purchase and of keeping the necessary in process records of these additions.

OUT OF PRINT--Not obtainable through the regular market, since the publisher's stock is exhausted.

OUTPUT--The transmission of information from the computer to a device where the information may be examined and/or removed. The form of the output may be readable by humans or by machines. Examples of the former are typewritten or printed pages; examples of the latter are punched cards or magnetic tapes*. In the case of CRT* displays, the image on the screen may be photographed in order to remove it from the device.

OUTPUT REQUEST--A request by the user to have certain information presented to him. The information is taken from the records* located by the system in response to his search request*.

OVERDUES--The record keeping designation applied to the process of noting when some circulating material has been returned after the due date of its load period.

PAMPHLET BINDING--1. Binding in which the sheets are stapled. The term applies both to pamphlets and to magazines. 2. The manner in which pamphlets and magazines are bound as they come from the publisher; usually stapled. 3. A form of repair* in which material is stapled into a stiff cover.

PART--One of the subordinate portions into which a volume* has been divided by the publisher. It usually has a special title*, half title, or cover title, and may have separate or continuous pagination, foliation, or register, but it is included under the collective title page or cover title of the volume which is intended to contain it. It is distinguished from a fascicle* by being a unit rather that a temporary division of a unit.

PARTITION--A division of computer memory or some storage medium space into two or more non-overlapping segments.

PASSWORD--A set of characters which is entered by a user to demonstrate that he is authorized to access or alter information in a file*.

PERIODICAL--A serial* publication appearing or intended to appear indefinitely at regular or stated intervals, usually more frequently than annually, each issue of which normally contains separate articles, stories, or other writings. Newspapers disseminating general news, and the proceedings, papers, or other publications of corporate bodies, primarily related to their meetings are not included in this term.

PERIODICAL DEPARTMENT--1. The part of a library where current issues of periodicals* and other serials* are kept for reading. 2. The administrative unit in charge of handling periodicals, which may include ordering, receiving, preparation for binding, circulation, etc.

PLATING--The process of preparing and pasting bookplates in books.

PREPRINT--An impression printed in advance of regular publication, as of a periodical* article, part of a book, or paper presented to a conference.

PRINCIPAL FILE--The file* containing the records* to be searched as well as the indexes, dictionaries, and thesauri associated with the file.

PRINTER--A device which can print up to 133 characters per line as lines are transmitted from a computer. Lines are printed at the rate of 500 to 1000 lines per minute.

PRIVATE FILE--A file* which may only be accessed by persons designated by the file manager*.

PROCESS SLIP--A card or slip, sometimes a printed form, which accompanies a book through the Catalog Department*, acquiring on its way all the information and directions necessary for cataloging fully. Also called Catalog Slip, Cataloger's Slip, Cataloging Process Slip, Copy Slip, Work Slip.

PROMPT--An output*, generally very short, to a terminal* which indicates that the system is ready to accept the next request from the user.

PUBLIC FILE--A file* which may be accessed by anyone who has a terminal available and is authorized to use the system.

PUBLISHERS WEEKLEY--A serial* with a weekly listing of recent American trade publications. It is the book trade journal for the United States.

PUBLISHER'S SERIES--A series of books, not necessarily related in subject or treatment, issued by a publisher in uniform style and usually with a common series title, as Cambridge Edition, Everyman's Library. Also called Trade Series and Reprint Series.

PW--See Publishers Weekly*.

QUERY--A request for information from the system. One type of query is a search request*.

QUEUE--A waiting line made up of requests to be processed by the system.

RECALL-The record keeping designation used to request the return of library material which another user has requested or when it is needed for reserve.

RECORD--A group of data elements* that are stored together because they share some common relationship. For example, in a library card catalog there is at least one record (card) for each book.

RECOVERY--The procedure that must be followed to transform a backup file* and an historical file* into a principal file* which is identical in content to a principal file that has been damaged.

REFERENCE COLLECTION--A collection of books and other material in a library, useful for supplying information, kept together for convenience and generally not allowed to circulate.

REFERENCE DEPARTMENT--1. The part of a library in which its reference collection is kept for consultation. 2. The administrative unit in charge of the reference work of a library.

REFERENCE QUESTION--Any question requiring that a Reference Librarian consult a reference tool. As opposed to information question*.

RENEWAL--Recharging of books to the same borrower at expiration of the previous loan period.

REPAIR--The partial rehabilitation of a work or damaged book, the amount of work done being less than the minimum involved in rebinding. Includes such operations as restoring cover and reinforcing at joints.

REPORT GENERATOR--A collection of computer programs used to create intricate output* formats*.

REPRINT--A new printing, without material alteration, from new or original type or plates, as distinguished from copies made by typing or reproductions made by a mechanical or a photomechanical process. A textual reprint is one whose text follows exactly that of a particular edition.

RESEARCH LIBRARY--A library provided with specialized material, where exhaustive investigation can be carried on, in a particular field, as in a technological library, or in several fields, as in a university library emphasizing graduate level research.

RESERVE BOOKS--A designation applied to a collection of material specified by the instructor of an academic course to be placed on limited circulation.

RESERVE CIRCULATION--Circulation of a collection of reserve books limited to a specified period of time; usually 2, 8, 24, or 72 hours.

RESERVE PROCESSING--The process of searching, ordering, charging, creating course and author catalogs* or lists and shelf lists* and special shelving of books for reserve circulation*.

RESPONSE TIME--The elapsed time between submitting a request to the system and receiving the results. With an on-line system the response time is the period between typing the last character of the request and receiving the next prompt* at the terminal*.

RETRIEVAL--The process of locating and examining information in a file*.

RETRIEVAL REQUEST--A request to locate and present information which is in a file*. A retrieval request has two parts, a search request* and an output request*.

REVISER--A cataloger who checks and corrects work in process, such as the assignment of classification numbers* and preparation of catalog entries*. Also, a cataloger or senior assistant who revises filing in the main catalog.

SDI--Selective dissemination of information. A system in which the output* accumulated by standing search requests* is periodically distributed to the requestors.

SEARCH--The process of locating in a file* information which meets the criteria* specified in a request.

SEARCH PROCEDURE--The plan of search used in a particular library or class of libraries, especially for certain types or categories of information searches.

SEARCH REQUEST--The specification of the criteria* which describe the record that a user wishes to see.

"SEE ALSO" REFERENCE--A direction in a catalog* from a term or name under which entries* are listed to another term or name under which additional or allied information may be found.

"SEE" REFERENCE--A direction in a catalog* from a term or name under which no entries* are listed to a term or name under which entries are listed. Other terms used are: "See" Cross Reference, "See" Subject Reference, "See" Card and "See" Reference Card.

SEQUENTIAL FILE--A file* which is ordered in a single sequence. In order to access any record in it, all preceding records must be passed over.

SERIAL--A publication issued in successive parts bearing numerical or chronological designations and intended to continue indefinitely. Serials include periodicals*, newspapers, annuals (reports, yearbooks, etc.), the journals, memoirs, proceedings, transactions, etc., of societies, and numbered monographic series*.

SERIAL CHECK-IN--The process of record keeping in which the receipt of serials* is recorded.

SERIAL HOLDINGS--A list of serials* held in a collection including volume*,and part*, and date information.

SERIAL RECORD--A record of the serial holdings* of a library.

SERIALS DEPARTMENT--The administrative unit in charge of handling serials*, which may include ordering, checking, claiming, cataloging*, preparation for binding, etc.

SERIES--1. A number of separate works issued in successions and related to one another by the fact that each bears a collective title generally appearing at the head of the title page, on the half title, or on the cover, normally issued by the same publisher in a uniform style, frequently in a numerical sequence. Often termed "monographic series," "monograph series," 2. Each of two or more volumes of essays, lectures, articles, or other writings, similar in character and issued in sequence. 3. A separately numbered sequence of volumes within a series or serial*.

SET--A series* associated by common authorship or publication. Specifically, a collection of books forming a unit, as the works of one author* issued in uniform style, a file of periodicals*, related works on a particular subject, or unrelated books printed uniformly and intended to be sold as a group; as, a set of Dickens; a set of works on sociology.

SHELF LIST--A record of the books in a library arranged in call number* order, nominally in the order in which the books stand on the shelves, hence the name.

SHARED CATALOGING--See NPAC*.

SIMPLE REQUEST--A search request* consisting of the name of one data element* and a value for the element. For example, LOCATE AUTHOR HARRISON.

SINGLE ACCESS FILE--A manual or automated file* such as a purchase order file in which the only rapid access to a specific record* is by searching the data element* (e.g.purchase order number) which determines the fille sequence*. If a vendors name is known but not the purchase order number, the purchase order could not be found without a record-by-record search from the beginning to the end of the file.

SL--See Shelf List*.

SOFTWARE--The computer programs processed by computer hardware*.

SPACE ALLOCATION--See storage allocation*.

SPINE--That part of the cover or binding which conceals the sewed or bound edge of a book, usually bearing the title, and frequently the author.

SPIRES--The Stanford Physics (or Public) Information REtrieval System. Acronym for the information storage and retrieval project.

STANDING ORDER--A purchase order for a series* or terminal set* whereby the library automatically receives each new volume or title as it is published.

STANDING SEARCH REQUEST--A semi-permanent search request* that is used to retrieve* and output* documents meeting the search criteria* as the documents are added to a principal file*.

STORAGE ALLOCATION--Reservation of a portion of computer memory or magnetic disk* for certain classes of information.

SUBJECT CATALOGING--That phase of the process of cataloging* which concerns itself with the subject matter of books, hence, includes classification and the determination of subject headings*.

SUBJECT ENTRY--An entry* in a catalog* or a bibliography under a heading* that indicates the subject.

SUBJECT HEADING--See Subject Entry*.

SUBJECT INDEX--See keyword index*.

STANFORD UNIVERSITY LIBRARIES--An administrative unit headed by David C. Weber encompassing all libraries at Stanford University with the exception of the Coordinate Libraries*. See Appendix D for a complete list of the libraries at Stanford.

SUL--See Stanford University Libraries*.

SYSTEM FAILURE--An unanticipated malfunction of some part of the system. An occasional side-effect of system failures is destruction of or damage to the contents of files*.

SYSTEM PROGRAMMER--One of the people responsible for the design, development, and maintenance of the system.

TERMINAL--A point in the system at which data can either be entered or output. For instance, typewriters and CRT* devices may be terminals in an on-line computer system.

TERMINAL SET--A set* with a definite number of volumes, some of which a library may not yet possess. As opposed to a Serial*.

TEXT-EDITING--To rearrange textual data for machine input or output, that is, to delete, insert, or reposition data, symbols, or characters.

TITLE--1. In the broad sense, the name of a work, including any alternative title, subtitle, or other associated descriptive matter preceding the author*, edition, or imprint* statement on the title page. 2. In the narrow sense, the name of a work, exclusive of any alternative title, subtitle, or other associated descriptive matter on the title page. 3. In counting library material, one unique work, irrespective of the number of volumes* and/or copies of that work.

TITLE ENTRY--The record of a work in a catalog or a bibliography under the title*, generally beginning with the first word which is not an article. In a card catalog* a title entry* may be a main entry* or an added entry*.

TITLE II--Title II-C of the Higher Education Act of 1965. See NPAC*.

TITLE II CARDS--The catalog cards received from the Library of Congress by libraries participating in the NPAC* program.

TITLE II FILE--A file of Title II cards*.

TOPIC INDEX--See keyword index*.

TRACING--1. In the broad sense, any record of entries* or references that have been made in connection with the cataloging* of a particular work or publication, or with establishing* a particular heading*. In the narrow sense, the record on the main entry* of the additional headings under which the publication is represented in the catalog*.

TRANSACTIONS--Additions, deletions and modifications of information in a file*.

ULS--See Union List of Serials*.

UPDATE--To add, delete or modify information in a file*.

UNION CATALOG--1. An author or a subject catalog* of all the books, or a selection of books, in a group of libraries, covering books in all fields, or limited by subject or type of material; generally established by cooperative effort. 2. A central catalog.

UNION LIST OF SERIALS--A catalog*, in alphabetical author or title arrangement, of periodicals* to be found in the United States and Canada. Gives catalog description of each title and names of libraries which hold the periodical.

UNIT CARD--A basic catalog card*, in the form of a main entry*, which when duplicated may be used as a base for all other entries* for that work in the catalog by the addition of the appropriate headings*.

USER SERVICES--The activities and division of the library which directly serve the public, i.e., circulation* and reference.

VOLUME--1. In the bibliographical sense, a book distinguished from other books or from other major divisions of the same work by having its own inclusive title page, half title, cover title, or portfolio title, and usually independent pagination, foliation, or register. This major bibliographical unit may have been designated "part" by the publisher, or it may include various title pages or paginations. 2. In the material sense, all that is contained in one binding, or portfolio, etc., whether it be as originally issued or as bound after issue. The volume as a material unit may not coincide with the volume as a bibliographical unit. When a physical unit designated "part*" by the publisher is too large or too extensive to be bound with one or more others, it is called a volume in collation, but in contents and notes the publisher's designation is followed. 3. For library statistical purposes, any printed, typewritten, mimeographed, or processed work, bound or unbound, which has been cataloged and fully prepared for use. In connection with circulation*, the term volume applies to a pamphlet or a periodical* as well as to a book.

WANT LIST--1. A file recording books and other material which are to be purchased when funds are available, prices have been reduced, or publications are available. Also known as Waiting List, Want File, Possible Purchase File, Desiderata. 2. A list of books or other material that a library wishes to acquire by exchange*.

WEIGHTED SEARCH REQUEST--A form of compound search request* The simple search requests* making up the weighted search request are assigned scores. The search criteria* are satisfied when the scores for the satisfied simple search requests equal or exceed the score assigned to the weighted search request.

WYLBUR--An on-line system* provided by the Stanford Computation Center for the manipulation of files containing alphanumerical* information.

APPENDIX B

PRELIMINARY ANALYSIS PHASE METHODOLOGY

There are five major sequential tasks in the Preliminary Analysis Phase of System Development. These tasks are:

1. DETERMINATION OF THE GENERAL OPERATING REQUIREMENTS of the organization. These requirements are stated as objectives to be met, products to be produced, and services to be provided.

2. STUDY AND DOCUMENTATION OF CURRENT OPERATIONS. This task, called fact finding, involves interviews with all levels of operating personnel and results in the compilation of organizational, procedural, and statistical information. This information is used as a basis for determining the detailed operating requirements of the organization and for performing detailed analysis.

3. STATEMENT OF LIMITATIONS. From the analysis of current operations in relation to requirements a statement of limitations is derived. There are several kinds of limitations.

a. Requirements are met but not at the level of efficiency or service desired.

b. Requirements are not met at all.

c. Requirements are not met adequately with the system.

The objective here is to identify those areas which could benefit from computer support and manual improvement. The purpose of library system development is to produce tools to eliminate limitations or reduce their effect.

4. LONG RANGE SYSTEM SCOPE. Areas of critical need are determined by establishing priorities among limitations. A Long Range System Scope is created by looking at the total need against the constraints of time and cost. This scope communicates what needs to be done in system development and what needs to be researched. The objective of the scope is to state what is included in a system designed to deal with existing limitations.

5. FIRST IMPLEMENTATION SCOPE. It is not possible or desirable to deal with all areas of need or to develop all aspects of a system before implementing a part of the system. Aside from the constraints of time and money much of the design in the long term results from continued research and from statistical information that only computer

manipulation can generate. It is important in a first implementation to strive for an optimal integration of computer and manual resources so that the areas in most need of computer help are aided and means for further research are provided.

In order that these five tasks are executed efficiently certain standards are imposed which reflect the methodology. Because of voluminous data collected in fact finding, standards were set for the recording of information.

Basically, these standards are:

1. The use of standard symbols for the representation of processes in flow chart form.

2. The use of common forms for the representation of statistical data.

3. The use of special formats for the narrative description of files, documents, and processes.

The application of these standards can be seen by looking at Appendix C, DOCUMENTATION OF THE CURRENT LIBRARY SYSTEM.

The analysis of the current system in relation to requirements is summarized as follows. First, individual processes are evaluated and limitations are identified. For example,

1. Inefficiencies in the manual system due to constraints such as:

    a. Single access files

    b. Proliferation of files

    c. Need for control information (e.g. activity reports)

    d. Need for improved forms quality and standardization

2. Frequently performed activities as candidates for computer support or manual improvement.

3. High cost areas if the cost can decrease based on computer support.

4. Bottlenecks in the processing flow.

Next the entire system is evaluated and the following are identified.

1. Duplication of effort (Processing and Files)

2. Inefficient work flow

3. Unnecessary processing

4. Inadequate information for decision making

5. The need for increased service or additional services

The long term scope is an overall approach for dealing with the limitations which have been identified. The plan for the first implementation scope is a choice which yields the highest return on investment and the best possible reduction of limitations in the short term.

APPENDIX C

Documentation of the Current Library System

(Sample pages only)

INTRODUCTION TO APPENDIX C (EXCERPTS)

Following are samples of five different forms used in the current system analysis included as Appendix C of the scope document:

1. Sample Flow Chart
   (Music Cataloging)
2. Sample Process Description Form
   (Music Cataloging)
3. Sample File Description Form
   (Shelf List-Music)
4. Sample Document Description Form
   (Card Set)
5. Sample User Services Survey Form
   (Math-Stat Library)

The full Appendix C is six volumes of analysis data:

Volume I.....Acquisition Department Flowcharts
Volume II....Acquisition Backup
Volume III...Catalog Flowcharts
Volume IV....Catalog Backup
Volume V.....Government Documents
Volume VI....User Services

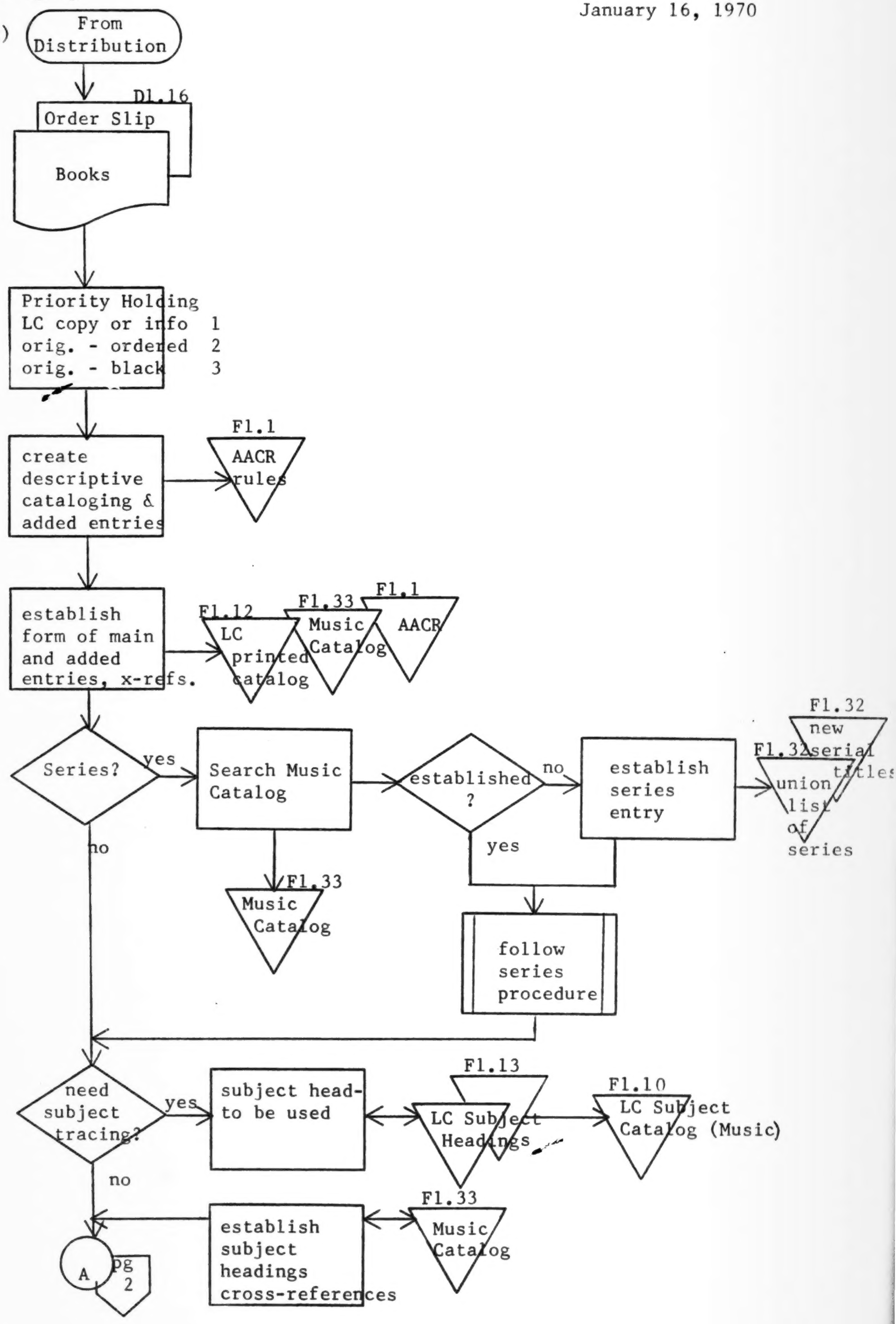
This documentation updates the 1967 System Study. It will be used as a basis for further study in the Detailed Analysis phase.

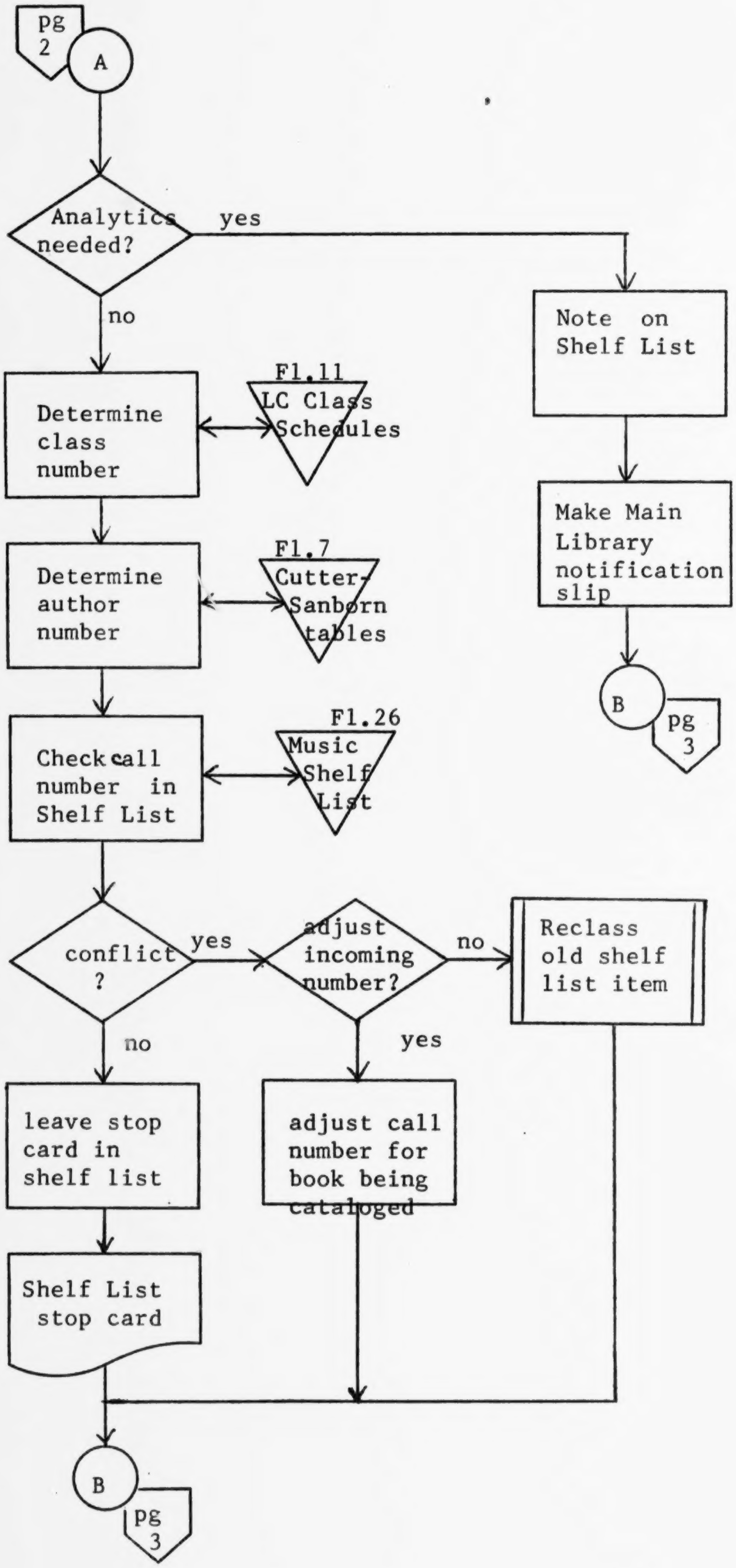
The backup volumes contain data on files, processes and forms. The flowchart volumes also contain data on personnel expenditures and organization.

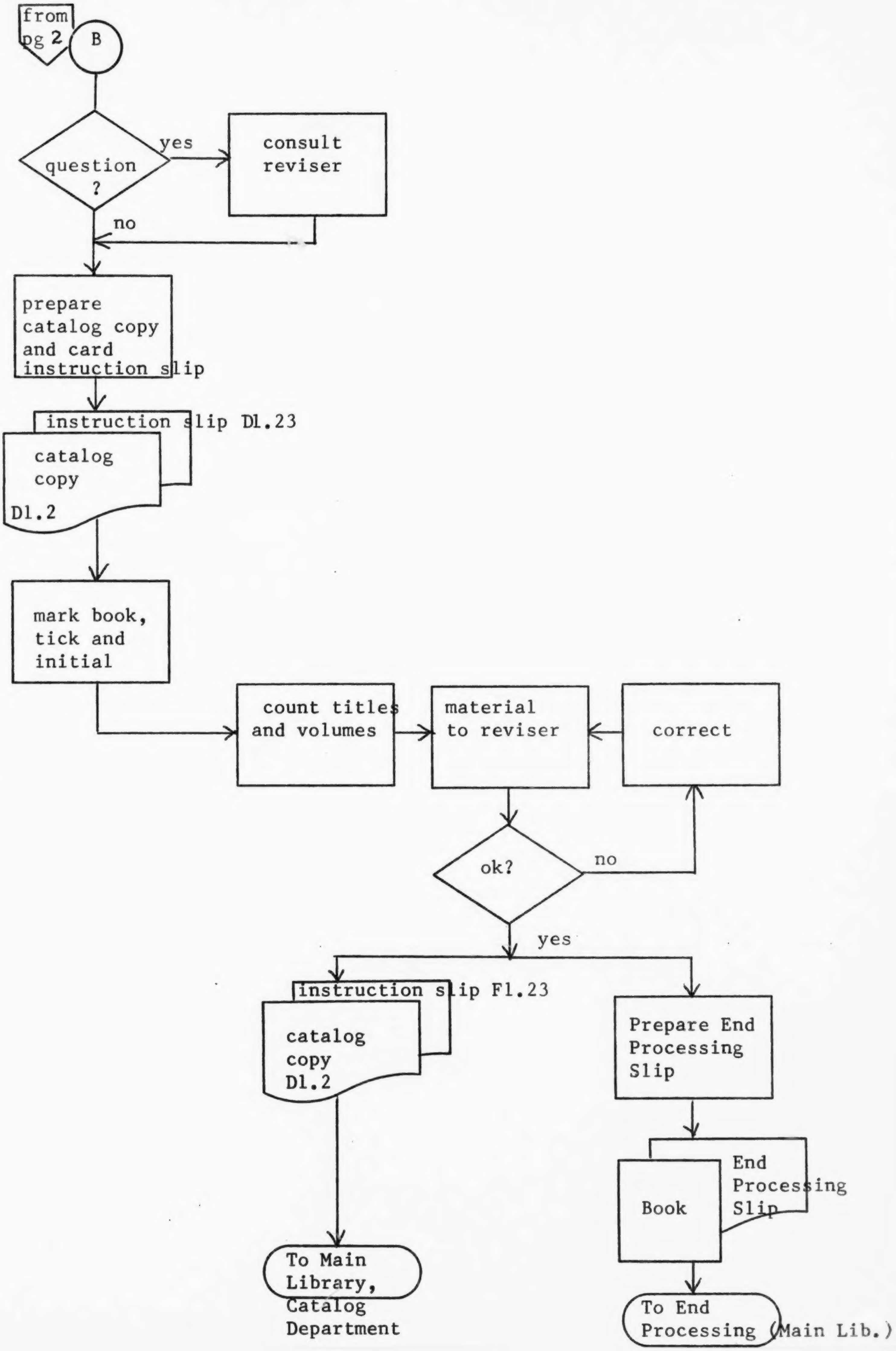
Files, forms and processes each have a unique number which is shown on the sample analysis forms. Certain sections of the forms (e.g. File Usage Characteristics of the file description form) will be completed during the Detailed Analysis Phase. Sets of this documentation are maintained in the respective departments and updated as changes occur. It is anticipated that this material will be used for continuous training and orientation of new personnel and in the analysis of procedures.

CATALOG DEPARTMENT
Music
Cataloging
(Original, LC
Card & LC Book)

Page 1 of 3
January 16, 1970


From Distribution
D1.16
Order Slip
Books
Priority Holding
LC copy or info 1
orig. - ordered 2
orig. - black 3
create descriptive cataloging & added entries
F1.1
AACR rules
establish form of main and added entries, x-refs.
F1.12
LC printed catalog
F1.33
Music Catalog
F1.1
AACR
Series?
yes
no
Search Music Catalog
F1.33
Music Catalog
established ?
no
yes
establish series entry
F1.32
new serial titles
F1.32
union list of series
follow series procedure
need subject tracing?
yes
no
subject head- to be used
F1.13
LC Subject Headings
F1.10
LC Subject Catalog (Music)
establish subject headings cross-references
F1.33
Music Catalog
A
pg 2


pg 2
A
Analytics needed?
yes
no
Note on Shelf List
Make Main Library notification slip
B
pg 3
Determine class number
F1.11
LC Class Schedules
Determine author number
F1.7
Cutter-Sanborn tables
Check call number in Shelf List
F1.26
Music Shelf List
conflict ?
yes
adjust incoming number?
no
Reclass old shelf list item
no
yes
leave stop card in shelf list
adjust call number for book being cataloged
Shelf List stop card
B
pg 3


from pg 2
B
question ?
yes
consult reviser
no
prepare catalog copy and card instruction slip
instruction slip Dl.23
catalog copy
Dl.2
mark book, tick and initial
count titles and volumes
material to reviser
correct
ok?
no
yes
instruction slip Fl.23
catalog copy
Dl.2
Prepare End Processing Slip
Book
End Processing Slip
To Main Library, Catalog Department
To End Processing (Main Lib.)

PROCESS NO.

P1.13

Stanford University Libraries
Automation Division

PROCESS DESCRIPTION FORM

1. Process name: Music Cataloging

2. Inputs: Books
Order Slip (D1.16)

3. Outputs: Instruction Slip (D1.23)
Catalog Copy (D1.2)
End Processing Slip
Books
Library Notification Slip

4. Files used: AACR Rules (F1.1)
Music Catalog (F1.33)
L.C. Printed Catalog (F1.12)
Union List of Serials (F1.32)
New Serial Titles (F1.32)
L.C. Subject Headings (F1.13)
L.C. Subject Catalog (Music) (F1.10)
L.C. Class Schedules (F1.11)
Cutter-Sanborn Tables (F1.7)
Music Shelf List (F1.26)

5. Personnel:

2.0 Professional
1.0 Library Assistants

6. Description:

Music Cataloging is done in the Music Library by Catalog Department personnel on permanent assignment to the Music Library.

In 1968/69, 2721 new titles were cataloged and 13,067 cards were filed in the Music Catalog and Shelf List (including the archive of recorded sound).

Analyst EAM
Date 1/10/70

AD-1 (12/69)

1. File Number 105

Stanford University Libraries
Automation Division

F1.26

FILE DESCRIPTION FORM

2. FILE NAME: Shelf List (Music)

3. LOCATION: Music Library

4. FILE CONTENT CHARACTERISTICS

a. File Sequence: Call Number

b. File Size: No. of Records: Av. ____

Present size 18,720 Growth Rate 2,232/yr.

c. File Content: Record description: Catalog Cards

Record size: Av. No. of Char ____ Max. ____ Min. ____

Record Retention Period Indefinite unless title cancelled or transferred.

5. FILE USAGE CHARACTERISTICS

a.	Process Name	Cataloging	Added cop/vol	Filing	TOTALS
b.	Process Number				
	Record Activity	Vol./Freq.	Vol./Freq.	Vol./Freq.	
c.	Records Input	____/hour ____/day ____/week 186 /month peak ____	____/hour ____/day ____/week ____/month peak ____	____/hour ____/day ____/week 186 /month peak ____	____/hour ____/day ____/week 372 /month peak ____
d.	Records Changed	____/hour ____/day ____/week 10 /month peak ____	____/hour ____/day ____/week 58 /month peak ____	____/hour ____/day ____/week ____/month peak ____	____/hour ____/day ____/week 68 /month peak ____
e.	Records Purged	____/hour ____/day ____/week 186 /month peak ____	____/hour ____/day ____/week ____/month peak ____	____/hour ____/day ____/week ____/month peak ____	____/hour ____/day ____/week 186 /month peak ____
f.	File Searches	____/hour ____/day ____/week 186 /month peak ____	____/hour ____/day ____/week ____/month peak ____	____/hour ____/day ____/week ____/month peak ____	____/hour ____/day ____/week 186 /month peak ____

Analyst GHB

Date 12 Jan. 1970

AD-3 (12/69)

1. DOCUMENT NUMBER: D1.2

Stanford University Libraries
Automation Division

DOCUMENT DESCRIPTION FORM

2. DOCUMENT NAME: Card Set
3. DOCUMENT NUMBER:
4. NARRATIVE DESCRIPTION:

Consists of the main entry card and all added entries, subject entries, shelf list cards and other miscellaneous cards as traced on the main entry card. Also includes any cross reference cards made to headings within the set when the book was cataloged.

5. DOCUMENT CREATION:

a) Created by	b) Volumes			
Department/Function/Process	Average	Peak	Frequency	Peak Period
Catalog Department - Cataloging	748,018/yr.		Daily	None

☒ Single - Each card is separate.

☐ Original of Multi part document

☐ Carbon of Multi part document

No. of copies	
Part No. ____ of ____	Original Document Name

6. DOCUMENT USE:

a) Processed by	b) Disposition
Department/Function/Process	Disposition (filed in, sent to, etc.)
Catalog Department	Sent for filing in Stanford files, to the Library of Congress and other locations. Sent for refiling into Stanford files as a result of added copy/added volume, recataloging or catalog maintenance operations.

AD-2 (12/69)

Analyst RCP
Date 1/14/70

DOCUMENT NUMBER.
91...2

DOCUMENT NAME: CARD SET

O acôrdo impossível.

Prado, Antônio Lázaro de Almeida.

PAVESE, CESARE, 1908-1950.

Prado, Antônio Lázaro de Almeida.

Almeida Prado, Antônio Lázaro de
see
Prado, Antônio Lázaro de Almeida.

Prado, Antônio Lázaro de Almeida.

PQ4835
A846Z77

O acôrdo impossível; ensaio sôbre a forma interna e sôbre a forma externa na obra de Cesare Pavese. Assis, Faculdade de Filosofia, Ciências e Letras, 1966.

205 p. 23 cm. port. NCr $5.50

(BB 68-

(LACAP 68-

Bibliography: p. 199-204.

1908-1950
1. ✓Pavese, Cesare. I. Title. NUC LA

✓PQ4835.A846Z77 858'.9'1209 68-911

Library of Congress [13]

Prado, Antônio Lázaro de Almeida.

PQ4835
A846Z77

Stack

SEL

Prado. Antônio Lázaro de Almeida.

PQ4835
A846Z77

Stack

Shelf List

Prado, Antônio Lázaro de Almeida.

PQ4835
A846Z77

O acôrdo impossível; ensaio sôbre a forma interna e sôbre a forma externa na obra de Cesare Pavese. Assis, Faculdade de Filosofia, Ciências e Letras, 1966.

Stack

205 p. 23 cm. port. NCr$5.50

(BB 68-5)

(LACAP 68-0469)

Bibliography: p. 199-204.

1908-1950

1. Pavese, Cesare. I. Title. NUC LA

PQ4835.A846Z77 858'.9'1209 68-91138

Official Shelf List

Library of Congress [12]

7

Prado, Antônio Lázaro de Almeida.

PQ4835
A846Z77

O acôrdo impossível; ensaio sôbre a forma interna e sôbre a forma externa na obra de Cesare Pavese. Assis, Faculdade de Filosofia, Ciências e Letras, 1966.

205 p. 23 cm. port. NCr $5.50

(BB 68-5)

(LACAP 68-0469)

Bibliography: p. 199-204.

1908-1950

1. Pavese, Cesare. I. Title. NUC LA

PQ4835.A846Z77 858'.9'1209 68-91138

Library of Congress [12]

Prado, Antônio Lázaro de Almeida.

PQ4835
A846Z77
[illegible]

O acôrdo impossível; ensaio sôbre a forma interna e sôbre a forma externa na obra de Cesare Pavese. Assis, Faculdade de Filosofia, Ciências e Letras, 1966.

205 p. 23 cm. port. NCr $5.50

(BB 68-5)

(LACAP 68-0469)

Bibliography: p. 199-204.

1908-1950

1. Pavese, Cesare. I. Title. NUC LA

PQ4835.A846Z77 858'.9'1209 68-91138

Library of Congress [12]

STANFORD UNIVERSITY LIBRARIES AUTOMATION DEPARTMENT

USER SERVICES SURVEY
JANUARY, 1970

A. General Information

1. Service point: Math - Stat Library
2. Number of volumes in collection: 19,902
3. Number of full time positions: 4

B. General Circulation Information

1. Number of charges 68/69: 15,841
2. Distribution of charges:
   - a. morning: 40
   - b. afternoon: 30
   - c. evening: 10
   - d. maximum/hour: 15
3. Number of overdues 68/69: 5/day
4. Number of holds/recalls 68/69: 25/day
5. Number of bills for lost/not returned books 68/69: 4/quarter
6. Number of delinquent bills 68/69:

C. Reserve Circulation Information

1. Number of volumes placed on reserve 68/69: 260/quarter
2. Number of reserve charges 68/69: 6,520
3. Distribution of charges:
   - a. morning: 12
   - b. afternoon: 10
   - c. evening: 6
   - d. maximum/hour: 8

D. File Information

1. Cumulative files:

	FILE	CURRENT SIZE
a.	catalog(s)	41,300
b.	shelf list	10,700
c.	serial holdings	523
d.	serial check-in	325
e.	other	
	index to uncat	
	material	3,000

2. Dynamic files:

	FILE	AVERAGE SIZE	MAXIMUM SIZE
a.	general charge	1100	1400
b.	reserve charge	12	20
c.	reserve shelf		
	list		
d.	reserve course	30 sheets	40 sheets
e.	reserve author		
f.	other		

E. Personnel Information

1. Number of positions in full time equivalents: 4

2. Personnel Distribution:

Number	Position	Ref	Circ	Paging Reading Shelving	Admin	Other	$/9mo
1	Librarian III	40%	5%		40%	15%	****
1	Library Specialist	40%	15%		5%	40%	****
1	Library Services Assistant	5%	70%	15%		10%	****
1	Student Assistan	5%	5%	70%		20%	****
TOTAL POSITIONS		.9	.95	.85	.45	.85	
TOTAL $							18,315

APPENDIX D

THE LIBRARIES OF STANFORD UNIVERSITY

- STANFORD UNIVERSITY LIBRARIES
  - BIBLIOGRAPHIC OPERATIONS
    - Acquisitions Department
      - Binding and Finishing Division
      - Exchange Division
      - Gift Division
      - Order Division
      - Serial Records Division
    - Catalog Department
      - Catalog Production and Maintenance
      - Meyer and Overseas Division
      - Monograph Division
      - Serial Division
      - Special Collections Division
      - Special Materials Division
  - CENTRAL SERVICES
    - Circulation Department
    - Financial Office
    - General Reference Department
      - Current Periodicals Service
      - Reference Desk
      - Asian Languages Library
      - Briggs Library
      - Classics Library
      - Communications Library
      - Graduate Program in the Humanities
      - Memorial Church Library
      - Modern European Languages Library
      - Tanner Philosophy Library
      - West Political Science Library
    - Government Document Department
      - Federal Documents
      - Foreign Documents
      - International Documents
      - State Documents
      - U.S. Classification
      - Microtext and Newspapers
    - Special Collections Department
      - Institute of American History
      - Jones Library
      - University Archives
  - UNDERGRADUATE AND BRANCH LIBRARY SERVICES
    - Art and Architecture Library
    - Cubberley Education Library
      - Main Branch

- Women's P.E. Library
- Meyer Memorial Library
  - Audio Division
    - Audio Services
  - Circulation Division
  - Reference Division
- Music Library
  - Music Library
  - Archive of Recorded Sound
- Science Department
  - Branner Geology Library
  - Computer Science Library
  - Dudley Herbarium Library
  - Engineering Library
    - Main Branch
    - Electrical Engineering/Solid State
    - Engineering Economic Planning Library
    - Guggenheim Aeronautics/Radio Science
    - Ryan Nuclear Technology Library
  - Falconer Biology Library
    - Main Branch
    - Systematic Biology Library
  - Math-Stat Library
  - Physics Library
    - Main Branch
    - Hansen Microwave Lab Library
    - Plasma Physics Library
  - Swain Chemistry Library
    - Main Branch
    - Chemical Engineering Library
- Hopkins Marine Station Library
- Inter-Library Loan
- Technical Information Service

- COORDINATE LIBRARIES
  - Food Research Institute Library
  - Hoover Institution Library
  - Jackson Library of Business
  - Lane Medical Library
    - Main Branch
    - Anatomy Library
    - Medical Microbiology Library
  - Law Library
  - Stanford Linear Accelerator Center Library

APPENDIX E

THE STANFORD LAW LIBRARY -- A POTENTIAL BALLOTS AND SPIRES USER

I. Legal Information Retrieval Overview
II. Proposal for an International Legal Studies Data Collection
III. International Legal Studies Searching at Another Law Library
IV. Library Automation and the Stanford Law Library
V. Conclusion

I. Legal Information Retrieval Overview

Law is one of the fields most heavily dependent upon libraries. The lawyer, judge, legislator, student and teacher constantly need to determine how problems have been resolved in the past, what rules have been established to solve new problems, and where gaps or inconsistencies exist in the legal structure.

No one lawyer can afford to maintain a library large enough to answer all his questions. The typical lawyer visits his local law library an average of twice a week. Many companies exist to supply law libraries with indexes, digests, annotations, and citation indexes. Law schools hire their professors primarily from the staffs of law reviews. One of the skills essential to law review work is expertise at legal research.

Numerous attempts have been made to introduce computer-based information retrieval systems into the legal research field. Only one commercial system has met with any success. The Aspen System Corporation has established and every month updates keyword indexes to statutory material. They sell research services primarily to state legislatures.

There are a number of characteristics which have helped to make the Aspen project successful: (1) legislative material has never been indexed on a national basis, (2) because of the terse and impersonal language of statutes they lend themselves to full text analysis, (3) contacts were established with state legislatures before beginning the project, (4) legislators historically entrust their research to others, and (5) legislators need complete and accurate research carried out many times a month.

Computer services which have merely duplicated the work of already existing manual systems have been unable to compete because of the high cost of computers. Systems that have relied on manual indexing have been criticized because the indexing was done by poorly paid, untrained or incompetent indexers. Many projects have been undertaken by one individual and abandoned when that individual lost interest. The professor who does his own research in his field of interest generally likes to do research and is not willing to entrust it to a computer. The lawyer who only occasionally does research tends to forget how to use a computer-based system between uses.

Before encouraging the creation of a legal information file the following points must be established: (a) that it fills a real need, (b) that computer operations are economical and (c) that the data base is created and used by more than one person.

For further information concerning legal research habits and legal information retrieval systems see "Research Habits of Lawyers", Morris L. Cohen, Jurimetics Journal, Vol 9, No 4, June 1969, pp. 183- 194 and "Legal Information Retrieval", Aviezri S. Fraenkel, Advances in Computers, Vol 9, 1968.

II. Proposal for an International Legal Studies Data Collection

The following is a resume of an interview with Prof. J. Myron Jacobstein of the Stanford Law School. Because of his interest in establishing a computerized data collection, his comments warrant serious consideration in the Current System Development Process.

1. At the present time it is extremely difficult to carry out comprehensive searches of periodical literature in the field of international legal studies.

2. One of the reasons for the difficulty is that there has been an explosive growth of literature in this field. No publishing company has done an adequate or thorough job of indexing and few law libraries can afford to hire indexers with expertise in international legal studies.

3. The Stanford Law Library would like to hire an expert indexer. If the results of his work could be accessed by or sent to other law libraries, then it would be easier to find funds to support such specialized indexing.

4. The indexer would have a list of descriptive terms. He would go through new acquisitions article-by-article and chapter-by-chapter preparing items to be added to the collection. It is estimated that he would prepare between fifty and one hundred items per month.

5. An item would consist of the author or authors,

date, title, journal or book in which the article appeared, and a list of descriptive terms.

Citations to prior acticles would not be included. A typical item would be assigned from five to ten descriptive terms.

6. It is not anticipated that bibliographic items would be available in machine readable form for input from other law libraries or from governmental agencies.

7. It is anticipated that once an item had been entered into the data collection that it would not be modified.

8. An author index and descriptive word index would be maintained provided that searches could be limited by date.

9. Professor Jacobstein anticipates that most searching would not be done by the researcher but by the law library staff. The reason for this is that the librarian is more likely to know what to ask for and how to use the system.

10. It would rarely be necessary to have rapid turnaround. If the researcher left his request with the librarian, the librarian could carry out all searches at some assigned time and present the results to the researcher the next day.

11. The actual data collection search should be iterative so that the librarian could interact with the collection, rephrasing the request when necessary, in order to minimize iterations involving the researcher.

12. When searching the data collection, the library staff would select terms from the same thesaurus (manually constructed) as used by the indexer.

13. The search language presently used by the system is adequate for legal searching. A typical search request might be: INTERNATIONALIZATION AND (CANAL OR WATERWAY) AFTER 1963.

14. It would be nice to permit standing search requests, but it is doubtful if this is economically feasible.

15. Any technique, such as attaching the data collection to the computer only at certain times in order to minimize costs, would be appreciated.

16. Since search requests would be exhaustive rather than cursory, it is expected that an average search would result in at least ten items retrieved.

17. There is no need for automated thesauri or synonym dictionaries.

18. Professor Jacobstein does not feel that he can estimate how many researchers would submit requests or how often they would do so.

19. He anticipates that after a number of years, the Stanford Law Library would receive search requests from the Pacific coast region of the United States.

III. International Legal Studies Searching at Another Law Library

The following is a resume of an interview with Thomas Reynolds who is in charge of services at the UC Berkeley Law Library. Since the University of California Law School at Berkeley specializes in international legal studies, it was felt that the library staff might be aware of the research habits of the law school staff in this area.

1. The UC Berkeley Law Library is different from the Stanford Law Library in that the library staff does not do research for the professors. Consequently, Mr. Reynolds could only say that he had not heard comments about any difficulties in researching international law questions.

2. When asked whether researchers at the law school would use the Stanford data collection proposed by Mr. Jacobstein, he could only think of two professors who might.

3. The UC Berkeley law library does not have anyone on its staff with special competence in the filed of international legal studies.

4. When Thomas H. Martin (SPIRES/BALLOTS Project) was a law student at Berkeley, he found international law to be the most difficult field in which to do research. The reason for this was not only poor indexing, but also that the source materials were spread between many libraries, each with its own indexing scheme.

IV. Library Automation and the Stanford Law Library

The following is a resume of an interview with Prof. J. Myron Jacobstein of the Stanford Law Library. Since the law library has shown an interest in using the in-process file of Project BALLOTS, it was felt that the needs of the law library should be considered in the next version of SPIRES/BALLOTS.

1. The law library would be willing to turn over

purchasing of books to the Stanford Libraries.

2. The law library, when ordering a book, would send whatever information they had concerning the book to some central office.

3. The book would hopefully be delivered directly to the law library. The library would then notify the central office that the book had arrived.

4. Prof. Jacobstein would like to be able to find out at any time the amount of money he had left in his purchasing account.

5. He would like to be able to discover before ordering a book whether or not one of the other libraries had ordered the book.

6. Periodically he would like to know what books had been ordered and not delivered.

7. Periodically he would like to know the amount of business conducted with each vendor during the period and the performance statistics of the vendor.

8. He would like to receive at least the Library of Congress card number or, if possible, a copy of the catalog card for books ordered by the law library and cataloged by the Library of Congress.

9. For books that are ordered but not yet cataloged by the Library of Congress, he would like to be notified as soon as the Library of Congress has cataloged the book.

V. Conclusion

In conclusion, there have been a number of attempts to create legal information retrieval systems. It appears that the Stanford Law Library might become a user of both the Generalized Information Storage and Retrieval System and the Library Automation System. None of their needs have been overlooked in the SPIRES/BALLOTS scope.

APPENDIX F

STANFORD LINEAR ACCELERATOR CENTER PARTICIPATION IN SPIRES

By Louise Addis

(SLAC Library)

The special characteristics of SPIRES as a Physics Information Retrieval Project were outlined by E.B. Parker in the 1967 SPIRES ANNUAL REPORT, as follows:

> "Five features characterize the SPIRES project and serve to distinguish it from other on-line information retrieval projects. The first is the strong behavioral science emphasis . . .
>
> The second distinguishing feature is the data base to be used in the system. The first criterion for selecting the data base is to be responsive to user needs, finding out user priorities rather than starting with assumptions that may not apply locally. . .the second criterion ... is to take advantage of whatever data bases are available in machine readable form that may be of some value to our users. . .
>
> The third distinguishing feature of the SPIRES is its focus on the development of adequate computer systems software and applications programming. . .
>
> The fourth distinguishing feature can be stated negatively. There is no local manual indexing. It is felt that what manual indexing is done would, in the interests of standardization, be better left to the developing national systems rather than attempting to index at a local level. Instead the concern is with adapting to on-line retrieval whatever indexing procedures are available or can be made available, and with indexing that can be done by computer (e.g., using title words in conjunction with word stemming and synonym dictionary procedures and using citation indexing procedures)...
>
> The fifth distinguishing feature is the nature of the liaison with relevant library operations and library automation projects. The project has excellent liaison with the SLAC Library . . ."

In keeping with the basic philosophy of SPIRES, the needs and priorities of potential SLAC users were explored in a series of interviews with SLAC physicists. A summary of their response is found in the first SPIRES ANNUAL REPORT.

In accord with interview findings, high priorities were given to the following data bases:

1. SLAC preprint collection
2. Nuclear Science Abstracts
3. Journals (at that time it was thought that the T I P tapes would be available to SPIRES)
4. DESY High Energy Physics Index

The DESY INDEX was later moved up into second place as the excellence of its keyword indexing and the completeness of its coverage of the high-energy physics literature became evident. A sample data base of NSA was created but a full NSA data base was moved down on the priority list because of its size. NSA, with its interdisciplinary coverage, contains on the order of 50,000 entries/year (against the 9,000 entries/year of the specialized DESY INDEX). Journal tapes have not yet been available at a reasonable cost; however, the high-energy physics journals are thoroughly covered in the DESY TAPES.

We believed then and still do that the SLAC PREPRINT COLLECTION plus the DESY INDEX would most closely meet the goals of providing a specialized user population (SLAC and Stanford high-energy physicists) with access to:

1. The most timely information -- preprints.
2. A large enough specialized data base to permit exhaustive retrospective searches.

The choice of these two high-energy-physics data bases would allow comparison of the effectiveness of two types of subject search:

1. Title word, author, and citation searching in a file (preprints) in which no manual indexing had been done.
2. Keyword, title word, and author searching in a file (DESY) in which extensive professional keyword indexing was provided.

Citation search capability for the preprint data base was regarded as particularly important since no "manual" indexing was planned for that file. The presence of citations would allow another subject approach (in addition to title word) to preprints. Libraries on the Stanford Campus were already subscribing to the vast, interdisciplinary SCIENCE CITATION INDEX in its printed version (3,000,000 citations/year, approximately $1200/year).

In physics, the citation search has several utilities:

1. General subject searching.
2. Tracing the fate of a specific piece of work.
3. Checking on whether a particular author is doing work that others find useful. ("Publish or perish" is giving way to "be cited or be sunk".)

As more physicists discovered SCI's purpose and utility, we found ourselves struggling through more and more manual searches in its profoundly unsatisfactory pages (the print is submicroscopic, it is always far behind, and references are skeletal and must always be looked up again in a second source to locate titles). We welcomed the potential capacity of SPIRES to allow us easily to bring these citation searches up-to-date in our own preprint collection (a year or more ahead of the printed index).

Originally it had been planned to allow citation searching in the same detail as in the printed SCI (by author and all types of papers). This proved technically difficult and the input too time consuming. We therefore limited citation input to bona fide journal references which could be entered and searched as simply a CODEN (for journal title), a volume No., and a first page No. Since the references on preprints are frequently sloppy and inaccurate, and since they will eventually appear in the printed SCI, this compromise seems a reasonable one to make. It does, however, make it impossible to do citation searching on conference papers and on preprints, and, of course, we cannot do a citation search by author.

The ultimate SPIRES system, should, of course, allow for the inclusion of the complete SCIENCE CITATION INDEX... if only for the benefit of the Medical School where it is perhaps most heavily used in printed form.

Then, to reiterate, our goal as SLAC users was creation of data bases of the most timely material, and one large enough and complete enough (with a professional subject index) to allow a thorough search to be made on any high-energy-physics topic. The chosen materials were:

1. SLAC preprint collection (3,000 documents/year)
   Searches to be utilized:
   a. Author
   b. Title word
   c. Report number
   d. Citation
   e. Date

2. DESY HIGH ENERGY PHYSICS INDEX (9,000 documents/year)
   Searches to be utilized:
   a. Keyword (up to 23 assigned to each document)

b. Title word
c. Author
d. Date

Since March 1968 a data base containing the SLAC Preprint accessions has been regularly created and maintained (weekly as permitted by hardware and software development). Input has been via the 2741 terminal located in the SLAC Library.

This preprint data base currently contains bibliographic information and citations for some 6500 documents, including all the high-energy physics preprints received in the SLAC Library for the period March 1968 to the present. Approximately 1000 documents are reports, preprints, and translations produced by members of the SLAC staff. The annual cumulative list of SLAC publications is produced from the SPIRES data base by a batch program.

Specifications for the conversion of the DESY FILE to the SPIRES format were completed in June 1969 (see list of SLAC-SPIRES documents). Though the programming has been nearly completed for the conversion, the data base has not yet been created.

In late 1968, SLAC proposed to and received a special grant from the AEC to begin printing and mailing (under the sponsorship of the Division of Particles and Fields of the American Physical Society) a weekly list of preprints "Preprints in Particles and Fields (PPF)". PPF began publication in January 1969. Master copy for the list is produced each Thursday from the week's SPIRES input data set.

PPF is currently used by nearly 1600 high-energy physicists and preprint libraries in the Western Hemisphere (including SLAC). The results of a questionaire sent to subscribers indicates that PPF is a success among high-energy physicists. (One enthusiastic user described it as "the best thing to happen in physics information in 50 years"). A popular feature is the "Anti-preprint" list which lists when and where previously announced preprints are published. Though PPF is not an integral part of the SPIRES system but a byproduct (which we would produce anyway, though more laboriously, without SPIRES), the enthusiastic response of the wider high-energy physics user community to "even a listing" of preprints is significant.

USER EXPERIENCE -- SPRING 1969

The SPIRES search and the preprint data base were sufficiently developed by Spring 1969 to put to the test of actual physicist users. (At that time, SLAC had only 2 or 3 on-line terminals outside of the library whereas there are now 23 such terminals.)

About 1200 people are employed at SLAC. The SLAC Library has a staff of 11. The "user population" for a SPIRES (with only a high-energy physics data base) consists of some 90 Ph.D. high-energy physicists (including about 20 temporary visitors from other labs), 25 graduate students (Ph.D. candidates) and up to 8 members of the SLAC Library staff.

The two-mile linear electron accelerator itself is a scientific instrument used by experimental high-energy physicists to conduct their research. Theoretical high-energy physicists do not use the accelerator but concern themselves with explanation and prediction. Since a high-energy physics experiment on a large accelerator may cost in the $100,000 range to perform, it is essential that work not be duplicated or undertaken unnecessarily. Therefore, keeping up (with preprints) is essential to the high-energy physicist.

The physicist users are as a group:

a. Very busy, irregular in their working hours Experimentalists, for instance, must work all night sometimes. Theoreticians usually arrive around 10:00 am and frequently work at home.
b. Quick thinking and quick learning.
c. Familiar with computers and likely to have a typewriter terminal close by (there are 23 terminals now at SLAC).
d. Interested in any real help they can get in keeping abreast of the information explosion.

As a part of the campaign to attract users, Prof. E.B. Parker spoke at a seminar. Some 15 physicists asked the SLAC Library to conduct searches for them and probably another 15 experimented with the terminal search themselves (though that was hard to keep track of). Several expressed their opinions in writing to E.B. Parker (I've attached a few of these letters of which I received copies).

The results of the user experiments with SPIRES in April and May 1969 may be summarized as follows:

1. The quick search response time of SPIRES was universally admired and the slow printout on the terminal was found universally annoying.

2. The plans for CRT devices, the save, and off-line print capability were heartily endorsed. Once the search points have been determined, the user usually doesn't wish to have to wait for printout at the terminal. He'd like his secretary to printout a WYLBUR dataset or pick up some printout at the Comp Center. He'd also like to be able to "flip through" a lot of

entries as you are able to do on the CRT, and sometimes save a few entries in a file of his own.

3. Almost every search included one or more citation elements.

4. Since the preprint data base was the only one available, no comprehensive retrospective searching could be done on-line. Consequently, much supplementary manual searching (in the DESY INDEX) was done by the SLAC Library staff (resulting in a serious work overload) during this period. Users were pleased with the results and it seems obvious that were DESY available on-line and publicized, many information needs would be better met. (We don't have the staff time to offer this kind of manual search service to everyone who needs it now and physicists don't have the time to do manual searches themselves except under the most desperate circumstances).

5. The hours 8:15-9:30 a.m. were awkward ones for physicists. If only an hour or so of on-line SPIRES service were to be available, the late afternoon would be the best for physicists. Also, in many cases, an hour was not enough time to complete the listings for a particular set of searches though the searches themselves might have taken only a few minutes.
   A 24-hour day, 7-days a week availability would be the most popular. An 8-hour day, 5-days a week next. A 2-hour service during the 4:00-6:00 p.m. period next.

6. Physicists would still like to be able to save selected references in their own files, and several of them would like some form of SDI.

7. Many users mentioned the desirability of left and right truncation on all indexed elements.

An INTERIM SPIRES FOR SLAC USE:

The current version of SPIRES with the following improvements would provide SLAC with a fairly versatile on-line information retrieval system with which to gain user experience during the next 18 months, and one for which a case for some funding might be made to our budget department:

1. Completion of the Anti-PPF program (1/2 done) would save 10/15 hours per month of the preprint librarian's time and nearly that much of terminal time (while adding an undetermined amount of program running time).

2. Addition of the DESY DATA BASE would allow thorough retrospective searching on high-energy physics topics. (The implementation of No.3 below is, however, necessary to allow use of the DESY FILE). It would undoubtedly save many hours of reference librarian time and allow us to provide our users with a much more efficient subject search service. The experience which could be gained from physicists actually using a large file would be helpful in planning the future SPIRES. In connection with the DESY file, we need frequency statistics for keyword usages (per my memo of 7/22/69 to Jim Marsheck).

3. The addition of an off-line print capacity would render the use of the current SPIRES system economically feasible. Frequently the listing of 75-100 documents may be required after a search which took one minute. To be paying $9 to $16/minute for a terminal listing (as opposed to further searching) is simply not economically feasible... even in the case where several terminals are being used at one time (a rather complex scheduling feat). On-line search capacity is essential for setting up a given search. Ideally the search results should be stored in a WYLBUR data set and listed from the terminal later...but given the impossibility of this, print off-line is a satisfactory substitute.

The following additional improvements would be helpful but not essential for the interim SPIRES:

1. The addition of a message of the day to be set by the SLAC data manager for the preprint and DESY files, allowing a report to the user on the latest additions to the file, or any other relevant information. At present, the user has no easy way of knowing what material may have been added to the file since his last search.

2. Clean up of the "type own" display format to eliminate the print-out of unabbreviated element names. The user, who knows enough to choose the elements he wants printed out, can get by without any identifying tags for the sake of faster print-out.

3. The availability of a batch program which uses the "Anti-preprint" data sets to add publication notes (PBN) to entries in the data base. (Space has been dummied in as an NSP element with each preprint entry.) After a preprint has been published it is much more useful to the searcher to have a journal reference than a report number (which he must check in the card catalog to locate).

4. The elimination of duplicate entries within the DESY data base (this problem is described in detail in the DESY User Spec) and perhaps the linking of entries between the DESY and preprint files.

We envision the interim SPIRES as an on-demand system...the "demands" being made to the SLAC Library where search times could be scheduled for convenience to users and economy to the system. If the PREPRINT and DESY data bases were both available, with an off-line print capacity, we would publicize the subject search, encourage physicists to submit search questions and to use the system themselves during "up-time". We would also expect to prepare a few experimental user profiles (R.E. Taylor and B. Richter would like to be guinea pigs for such a project) to see if individualized lists of new high-energy physics documents could be sucessfully prepared using the search points available in these two files. Faculty members at CALTECH have also expressed interest in an arrangement allowing them to submit searches to SPIRES from time to time, probably via the SLAC Library.

THE ULTIMATE SPIRES

We envision the long-range SPIRES as a 24-hour/day, 7-day/week service, utilizing CRT, allowing individuals to create their own files, either from scratch or by copying out of larger data base files, and allowing users access to a spectrum of large special-subject data bases. A list of machine readable reference services most of which are currently available in printed form on the Stanford campus is attached to this document. (It would be interesting to poll the other science libraries, including Medicine to see which indexes they'd most like to have on-line).

It is, of course, essential that the cost to the user of the ultimate SPIRES be "reasonable."

INFORMATION RETRIEVAL

Certainly the ultimate SPIRES should be able to accommodate the SCIENCE CITATION INDEX as well as the more conventional indexes. At SLAC, we would hope for the eventual conclusion of the following large data bases:

1. PREPRINTS
2. DESY
3. NUCLEAR SCIENCE ABSTRACTS
4. SCIENCE CITATION INDEX (Physics and technology sections)
5. U.S. GOVERNMENT RESEARCH AND DEVELOPMENT REPORTS
6. STAR (NASA)
7. PHYSICS JOURNALS (AIP)
8. CHEMICAL ABSTRACTS (some subset of)
9. ENGINEERING INDEX (if available)

The first four of these are the most important to us.

It would seem reasonable that the ideal SPIRES be designed to accommodate any and all of the available machine readable records for which there were sufficient need among Stanford users.

The ultimate SPIRES also should allow the user or the user's "agent" such as the library, to maintain "profiles" of the user's information interests. These should be easily changeable, should be in the regular SPIRES search format (i.e. a Jones, J. and not a Smith, etc.) and should be automatically activated when new material is added to the file. Formating of the output from the profile searches will be very important since it must make very clear to the user which elements in his profile are producing "hits" and which are not.

Experience gained using a relatively large file during the interim SPIRES should be utilized in the design of the SDI features of the ultimate SPIRES. It would be desirable to draw heavily on the experience of the Lawrence Radiation Laboratory group using NSA for SDI experiments.

LIBRARY ROUTINES

Eventually, we should like to be able to "check in" the preprints received, on a SPIRES terminal rather than in our manually maintained file. We wish to "weed" with the aid of SPIRES instead of entirely manually as at present. (Now the preprint librarian personally compares the Tables of Contents of each new physics journal with our preprint holdings to locate published preprints.) Ultimately, we hope that a "weed list" can be prepared weekly by SPIRES from a comparison of new journal tapes with the preprint data base. The preprint librarian can check the "weed list" for mismatches. The preprint data base could then be updated (PBN added) and master copy for an anti-ppf be produced.

To eliminate double input, we need to produce catalog cards (or a cumulative book catalog) for our preprint collection. (We prefer catalog cards at present.) Ability to produce catalog cards from SPIRES input would allow us to consider conversion of our entire cataloging operation to "SPIRES". Conversion of our manual circulation system to an on-line (or batch) scheme might logically follow. (Currently, circulation files are maintained by call number and by borrowers names.)

EDP methods have been used for serials handling in the SLAC Library since 1963. At present all but two staff members participate more or less regularly in projects involving either keypunching or on-line data set creation. On the whole, attitudes are favorable toward further ventures into automation.

A POSSIBLE INDIRECT SLAC SUBSIDY TO SPIRES

The thorough exploration of the possibility of our using our own time-sharing system (CRBE) to create weekly preprint data sets which could then be transferred to the campus facility for incorporation into the SPIRES data base. I have explored this possibility enough to find that it is a good deal less convenient than our current system and might run aground on some technical difficulties. (data set size limits) Discussions are needed between a member of the SPIRES programming staff and the SLAC Computation Center, however, to determine whether it could indeed be done and how much programming would be needed to make it possible.

Moving the SLAC-SPIRES dataset creation to the SLAC computer would allow us to provide a large indirect subsidy to the SPIRES project without actual transfer of funds.

SLAC-SPIRES DOCUMENTS -- Formal and informal

A. INPUT FORMAT

1. Computer Note No. 30, INPUT FORMAT FOR SLAC PREPRINTS, LA, 28 Nov 1967.

An annotated version of this note is kept current (by hand) in the SLAC Library. (It needs to be reissued in a formal revision.)

2. COMMONLY USED CODEN

3. Title symbol conversion list and hyphenation conventions for physics preprints.

4. Brief Outline Guide to Wylbur for operator reference.

B. PREPRINTS IN PARTICLES AND FIELDS, a weekly newsletter in two parts

   1. PPF (the preprint announcement section)

      a. PREPRINTS IN PARTICLES AND FIELDS FORMAT SPECIFICATION, LA, Dec 1968.
      (Program was written by Ken Siberz, Jan 69, which creates master copy for PPF according to specification)

      b. PROCEDURES FOR USING PPF LIST CREATING PROGRAMS, LA, current.

      c. Time and length job records.

   2. ANTI-PPF (the section announcing publication of ex-preprints)

      a. SPECIFICATIONS FOR 'ANTI-PPF' LIST PRODUCING PROGRAM, LA, Oct 69.
      (Programming is not yet finished for this application.)

C. UPDATE

   1. CURRENT PROCEDURES FOR UPDATING THE PREPRINT DATABASE USING SLAC INPUT DATA SETS AND THE SPIRES PROGRAM.

   2. PROCEDURES FOR CHECKING THE BUILD AND HANDLING CORRECTIONS.

   3. TIME, AND LENGTH, AND JOB RECORDS.

D. SLAC PUBLICATIONS LIST

   1. USER SPEC FOR SLAC PUBLICATIONS LIST, LA, Dec 68.

      The SLAC Publications lists are an annually produced cumulative listing of all preprints, reports, translations, and internal reports done at SLAC.

LIST A -- is a cumulative listing of all SLAC preprints, reports, and translations currently. This amounts to about 1000 entries in the "Preprint Data Base" by author, Report No. and by subject. Master copy for list A has been produced twice and published since the programming was completed.

LIST B -- is a cumulative listing of all the SLAC internal reports (Technical notes) by author, Report No., and keyword.

LIST B has never been produced. The input dataset containing some 600 entries has been ready at SLAC since August 1969. It has never been added to the preprint data base... initially because of technical limitations on the size of the data base and currently because of uncertainty about the immediate future of the SLAC role in SPIRES.
The TN entries are the only ones which have actually had keywords assigned locally by the SLAC Library cataloger (using the DESY KEYWORD system).
We had hoped to have a data element level update available before committing the TN's to the data base since we would like to experiment with the effectiveness of the keywords and change them at will.

E. CATALOG CARDS

1. SPECIFICATIONS FOR USING SLAC INPUT DATASETS TO PRODUCE CATALOG CARDS, LA & KB, Aug 1968.

This card-producing specification with a few minor revisions is still valid for producing catalog cards for the SLAC Library catalog. A few decisions remain to be made -- the type of card to use...whether to produce cards on the 2741 terminal or on the line printer. . . how to handle the name authority list. At the present time we are doing "double input" as a part of participation in SPIRES...one staff member continues to make catalog cards (using a stencil and a cardmaster) while the terminal operator inputs the same information into a WYLBUR data set. Programming time has never become available for this application.

F. SEARCH

1. QUICK GUIDE TO SPIRES PREPRINT SEARCH, LA, Jun 69.

A summary of machine readable reference materials (available in 1968) extracted from C.P. Bourne, "Machine language bibliographic text and data records", Lecture Notes for University of Oregon 1968 Workshop on Library Mechanization.

Table I

Examples of Bibliographic Files Presently Distributed in Magnetic Tape Form

American Institute of Aeronautics and Astronautics - INTERNATIONAL AEROSPACE ABSTRACTS (IAA)

American Petroleum Institute - PETROLEUM ABSTRACTS

American Society for Metals - REVIEW OF METAL LITERATURE

Atomic Energy Commission - NUCLEAR SCIENCE ABSTRACTS

Chemical Abstracts Service - CHEMICAL ABSTRACTS CONDENSATES; BASIC JOURNAL ABSTRACTS; CHEMICAL-BIOLOGICAL ACTIVITIES; CHEMICAL TITLES; POLYMER SCIENCE & TECHNOLOGY

Clearinghouse for Federal Scientific & Technical Information - U.S. GOVERNMENT RESEARCH & DEVELOPMENT REPORTS

Derwent Publications, Ltd. - FARMDOC; PLASDOC; RINGDOC

Engineering Index, Inc. - Electrical/Electronics, and Plastics Sections of ENGINEERING INDEX; ENGINEERING INDEX MONTHLY

IFI/Plenum Data Corporation - UNITERM INDEX TO U.S. CHEMICAL & CHEMICALLY RELATED PATENTS

Institute for Scientific Information - ISI SOURCE DATA TAPES
ISI CITATION TAPES; INDEX CHEMICUS REGISTRY SYSTEM

Library of Congress, MARC Project - LC catalog records

NASA - SCIENTIFIC & TECHNICAL AEROSPACE REPORTS (STAR)

National Library of Medicine - MEDLARS tapes for INDEX MEDICUS

New York Times - NEW YORK TIMES INDEX

Pandex - PANDEX Airmail Weekly Tape Service
University of Tulsa - Indexes & Search Tapes to PETROLEUM ABSTRACTS

Table II

Examples of Available but Generally Non-distributed Machine Bibliographic Records

American Bibliographic Center - HISTORICAL ABSTRACTS; AMERICA: HISTORY AND LIFE

American Geological Institute/ Geological Society of America -
BIBLIOGRAPHY AND INDEX OF GEOLOGY EXCLUSIVE OF NORTH AMERICA

American Society for Information Science - DOCUMENTATION ABSTRACTS

Applied Mechanics Review - APPLIED MECHANICS REVIEW

BioSciences Information Service - All titles ever published by BIOLOGICAL ABSTRACTS,
BOTANICAL ABSTRACTS, &
ABSTRACTS OF BACTERIOLOGY

Compendium Publishers International Corporation -
SEARCH-DATA --
Marketing research information on chemicals and the chemical industry

Galton Institute - PERCEPTUAL COGNITIVE DEVELOPMENT

Educational Research Information Center - RESEARCH IN EDUCATION

National Agricultural Library - PESTICIDES DOCUMENTATION BULLETIN

National Library of Medicine - Current Catalog

Project URBANDOC - Bibliographic records related to urban planning & renewal

R.R. Bowker Company - PUBLISHER'S WEEKLY; FORTHCOMING BOOKS; PAPERBOUND BOOKS IN PRINT; SUBJECT GUIDE TO BOOKS IN PRINT; CHILDREN'S BOOKS FOR SCHOOLS AND LIBRARIES

U.S. Geological Survey - ABSTRACTS OF NORTH AMERICAN GEOLOGY
University Microfilms - DISSERTATION ABSTRACTS

Table III

Examples of Data Files Presently Distributed in Magnetic Tape Form

American Society for Hospital Pharmacists - Descriptive information and identification information for all major pharmaceutical products

Department of Commerce - 1958-1965 Industry Profiles (basic data relating to employment, payrolls, manhours, value of shipments value added by manufacturer, and capital expenditures for 409 manufacturing industries from the 1963 and 1965 Bureau of the Census Annual Survey of Manufacturers)

Dun & Bradstreet - Marketing facts on 5700 electronics manufacturers in the U.S. and Canada

Frost & Sullivan, Inc. - Defense Market Measures System (over 250,000 descriptions of U.S. Government contracts)

Investment Statistics Laboratory - Daily prices and volume of trading of all stocks on New York and American Stock Exchange since 1962

McGraw Hill - COMPUSTAT -- data on 1500 leading industrial and utility corporations

University of California at Los Angeles - Political Census File -- electoral and demographic records of Los Angeles County, including registration and voting records from the 1958, 1960, 1962, and 1964 general elections.

ATTACHMENT TO APPENDIX F

Excerpts from letters to E.B. Parker commenting on the SPIRES system as viewed by physicists.

Letter dated 7 April 1969 from H. Saal, Experimental Group C

> "I would like to take this opportunity to comment on the SPIRES system now operating at Stanford Linear Accelerator Center.
>
> I very much appreciate this existing facility, and look forward to its expansion and growth in the future. Particularly in the field of high energy physics, where selective access to large numbers of preprint data prior to formal publication is critical, such a tool is welcomed.
>
> Certain current limitations, such as the lack of uniform keywords, need to be overcome before the system can reach its full potential. I hope this effort will continue to be supported, and new features implemented in the manner. . .described to me."

Letter dated 21 April 1969 from D. Yount, Experimental Group D

> "This note is to express our appreciation for the work you and others have done in developing the SPIRES system.
>
> The streamer chamber group at SLAC is in the midst of a comprehensive article on meson photoproduction, and already we have used the SPIRES system to good advantage. Among the listings we have requested are: RHO Title Search (68 documents), RHO PHOTOPRODUCTION (13 documents), and articles referring to our own report, Phys. Rev. Letters 21, 841 (1968) (5 documents), which appeared some seven months ago. In each case, the lists have included the most recent and most inaccessible references, thus permitting a more thorough documentation than would otherwise be practical. We look forward to the expanded data base and increased flexibility that we understand are included in your future plans for the SPIRES system."

Letter dated 4 April 1969 from E.L. Garwin, Group leader, Physical Electronics

> "I have looked at the SPIRES information retrieval system which you have been developing, and am very enthusiastic about the potential of this kind of system to aid not only my own work but the work of applied physicists generally. Applied physicists have a particularly acute need for extensive and rapid bibliographic information services and should find your kind of interactive retrieval system very helpful.
>
> I am especially interested in the citation indexing capability demonstrated in the current SPIRES preprint data base. It is, for instance, a great time-saver for users to have titles and sources of citing articles instantly available.
>
> SPIRES will be most useful for my own work when it has a large collection of references, for example, a five-year accumulation of "Nuclear Science Abstracts," at least a two-year accumulation of the "Science Citation Index," and ideally, several years of "Chemical Abstracts."
>
> I hope you are able to obtain continued support for this important development effort."

Letter dated 9 May 1969 from S. Drell, Deputy Director, SLAC

> "I should like to congratulate you on the contribution which the development of the SPIRES system is making to the easing of the information crisis in science, particularly in high-energy physics, here at Stanford.
>
> The ever-growing flood of preprint and journal literature makes it essential for the physicist to have quick, direct access to the relevant literature of his field. He may then spend his time working rather than searching, confident that he is tackling something new rather than duplicating the old.
>
> The SPIRES concept of the comprehensive on-line search with output available on a CRT-scope should provide just such a mind-augmenting system for information retrieval. Even at its present stage of operation as a prototype system only, SPIRES shows great power and flexibility and has provided what I asked of it in connection with my own research efforts.
>
> The title work search combined with the citation search is an effective technique for exploring the high-energy

physics preprint collection which has been, until SPIRES, inaccessible by subject. Several years of DESY HIGH-ENERGY PHYSICS INDEX and NSA files would, of course, greatly enhance the value of the system for searching. The inclusion of extensive SCIENCE CITATION INDEX files would benefit not only physicists, but the whole campus scientific community.

I hope that SPIRES will continue its development along the lines presently proposed. Such a system has much to contribute to easing the flow of information and ideas in all fields."

Letter dated 10 May 1969 from Prof. A.H.Rosenfeld, Secretary, Division of Particles and Fields of the American Physical Society.

"Professor Panofsky and I want to thank you on behalf of the APS Division of Particles and Fields for the major contribution made by the SPIRES project to the success of our publication "Preprints in Particles and Fields (PPF).

As you know, we recently conducted a survey of our 1500 subscribers and received an overwhelmingly favorable response to PPF. Several physicists believe PPF to be the most useful advance in physics information in the last decade.

Also, I know that Si Pasternack, the Editor of the Physical Review is enthusiastic about the PPF way of dealing with the preprint problem and himself uses "Anti-preprints" extensively in editing the references in papers for the Phys. Rev. (Journal editors have in the past been in strong opposition to other more formal preprint handling schemes.) Of course all journals have this problem of updating references to preprints. . .

I understand that additional SPIRES efforts are plannned in connection with the "Anti-preprints" section. This will help in further easing the burden on SLAC Library personnel in the production of this bulletin which is such a boon to communication among high-energy physicists."

APPENDIX G

TUTORIAL: INFORMATION STORAGE AND RETRIEVAL

This appendix is intended to serve as an introduction to the concepts involved in the view of Information Storage and Retrieval held by the staff of the SPIRES/BALLOTS project. It is not a survey and does not attempt to cover all relevant problems or all of the techniques that have been developed in this area of computer technology.

A. TERMINOLOGY

In order to clarify the following introduction to the field of Information Storage and Retrieval, several key terms are defined. These terms are: files, retrieval, sequential files, direct access files, search and output. Other important terms are defined as they are introduced in the text.

A FILE is any body of information which exists on some storage medium and is structured so that segments of the information can be located and extracted in a systematic way. An example is a card catalog in a library. The storage medium is the cabinets containing cards and the systematic organization is an alphabetic ordering by author, title and subject. Another file, similar in structure though different in content, is the set of employee records stored in manila folders in a personnel office. A somewhat different kind of file is the multiple listing maintained by real estate sales firms. This file might be organized by price range, number of rooms or architectural style.

Once a file is established, the process of locating and extracting information is called RETRIEVAL. This process consists of several actions. The first is to formulate a QUERY, e.g., find the names of all books in the library pertaining to Serbian History. The second action is to look for relevant information. In this example, the inquirer scans the cards for the phrases 'Serbia-History' and 'History, Serbian'. The final action is to remove or copy the segments of information which satisfy the query conditions. In this example, removing the catalog cards, even momentarily, is not acceptable; therefore, the retriever would copy the information onto a loan request, charge slip or his own 3x5 cards.

Files are usually classified as SEQUENTIAL or DIRECT ACCESS although some might be considered a combination of the two. A SEQUENTIAL FILE is ordered in a single manner.

In order to locate any particular item of information, it is necessary to pass over to all preceding items.

In a DIRECT ACCESS FILE, any item may be retrieved without passing over a number of other items. To illustrate the difference, consider two files consisting of film representing a pictorial record of a vacation to Oregon. One of these files is a reel of 16 mm film and is a sequential file. To show Crater Lake, all of the scenes recorded prior to that must be passed over first. The second file is a set of 35 mm slides and represents a direct access file. To show the scenes of Crater Lake, only that specific set of slides need be projected. To locate the required set quickly, a list of scenes is maintained in some detail indicating which box or tray each set is stored in. This list is an index to the file. The concept of an index will be discussed later since it is central to the feasibility and utility of information storage and retrieval.

The process of locating the information described by a user in his query is called SEARCHING. The query is sometimes called a SEARCH REQUEST. The process of presenting the segments located by the search is called OUTPUT. Also, the resulting copy of the information is called the OUTPUT for the request. Both of these functions are discussed in later sections in more detail. Consider a search request applied to a personnel file to locate the records of all employees under 30 years of age earning in excess of ten thousand dollars. The computer, assuming a sequential file, examines the record of every employee in the file and checks the age and salary. This operation constitutes the search. For each record meeting the conditions specified in the query, the items of information in that record which were specified in the OUTPUT FORMAT (for instance, name, position and department) are printed. This is the output process for the example.

B. FILES

Files are stored on various media. Some of these are cards, sheets of paper, film and metal plates and are collected on shelves, in cabinets, in notebooks, on racks or in bound volumes. These files may contain many different kinds of information, as:

1. purely numeric items in a volume of statistical tables,
2. blueprints in an architect's file,
3. the textual content of an encyclopedia,
4. the mixed format of a personnel file.

The latter contains items which are numeric (age, salary), textual (references), coded (skill categories) and special forms (date of employment, inverted name).

Although most files not stored on computer equipment are sequential in nature, they usually have some of the characteristics of a direct access file. For example, an encyclopedia is organized by subject matter in alphabetical order. However, since each volume has the range of subjects printed on the spine, a person who is seeking information may narrow his search immediately to a specific volume. He then will find the correct page by making successive approximations and will have completed the entire search in a matter of seconds. The limitation of this technique is that the user of the encyclopedia must be familiar with the subject classification and often he does not retrieve all the relevant material. For instance, if he is looking for biographical material on Abraham Lincoln, he may not find the additional information contained under the subjects of Ulysses Grant or Appomattox.

Similarly, if a personnel file is ordered alphabetically on last name, it may be accessed quite efficiently when retrieving the records of individual employees whose name are known to the searcher. However, for any other type of retrieval, additional capability is required. This could be achieved by having multiple copies of the file, each of them ordered on some attribute of the employee, e.g., social security number, job classification, review date. Obviously, this would be too expensive and would lead to an unacceptably large number of errors. A more manageable alternative is to maintain a list for each category of information which INDEXES the file. For instance, a list could be maintained of all job classifications. Under each entry in this list would be a list of names of employees having that classification. If someone wished to send a memorandum to all executive secretaries, he could consult the list and obtain their names. From the file itself, he could get the company address for each.

The technique just described transforms an essentially sequential file into a form of direct access file. However, it is still somewhat cumbersome and prone to errors since, for each change in the file, one or more of the indexes may have to be changed. Another difficulty arises from the fact that the file exists in only one location while people in many locations may need to access it. Also, if one user of the file has removed a record, other users must wait until the record is returned. Many of the problems inherent in manual files can be resolved by placing them in the environment of a computerized information storage and retrieval system.

A sequential file to be accessed through a computer is normally stored on MAGNETIC TAPE. These tapes, and the mechanisms which write information on them (and read from them) are similar to home recorders, though larger, more complex and more expensive. A file on tape is purely sequential. It is restricted to a single ordering, and to access any one record, all previous records on the tape must be passed over. Another limitation of tape files arises from the fact that the tapes are normally stored OFF-LINE, i.e., on racks away from the computer. The information may be retrieved only when the tape is mounted on the read/write mechanism. Primarily because the tapes are stored off-line, this type of file is relatively inexpensive. It is a satisfactory mode of storage for files when the normal requirement is for large amounts of information on an infrequent basis rather than small amounts frequently and rapidly.

Computerized direct access files are normally stored on MAGNETIC DISKS. These disks are similar to phonograph records except that the recording is done magnetically rather than by physically cutting into the disk. The storage mechanism for direct access files is similar to the arm on an automatic changer. The disk access mechanism has the read/write cartridge on an arm which moves across the disk allowing rapid access to any track. Thus the information stored on a track of the disk may be accessed without reading over the information on other tracks. For instance, if each track held one employee record, then any employee record could be retrieved immediately if the numeric ADDRESS of the track for that employee were known. Having a sound method for determination of track addresses is one basis of a successful information storage and retrieval system of this type.

For the personnel file referred to above, retrieval requests will normally be stated in terms of employee attributes such as name, job classification, review date and skill categories. Other attributes such as home address and name of spouse are in the record of the employee but are not normally used in the formulation of queries. The attributes of the employees are called the DATA ELEMENTS of the file. The data elements which can be used in retrieval requests are called the ACCESS POINTS for the file. In a file of bibliographic references, the data elements would be items like author, title, publisher, number of pages and date of publication. The access points might be author, title and date of publication.

A means of creating access points for files is to construct an INDEX for each data element which is used for searching. The set of indexes is also stored on disks, in an order which allows efficient searching. An example is the AUTHOR INDEX for a bibliographic file. Assume that, on the average, the names of 50 authors can be stored on a single track of a disk and that the file contains the names of 2000 authors. The names are stored, in alphabetical order, over 40 tracks. In addition, a master track contains the first name on each track of the index. Each author's name has one or more addresses stored with it which indicate the location of each bibliographic reference associated with that author. If a user specifies the name Harrison H. Smedley in his search request, the following steps are taken by the computer. The master track for the author index is retrieved from a disk. The list of names in it is searched for a pair of consecutive names between which Smedley falls alphabetically. The address associated with the name which comes before Smedley is used to retrieve another track from the disk. If that track does not contain the name Smedley, the user is informed that the file has no references for Smedley. If, on the other hand, an entry for Smedley is found in that track of the index, the addresses contained in the entry allow the computer to retrieve all of the bibliographic references in the file for works authored by Smedley.

The organization of indexes in an information system is actually more complex than this but the general principle is the same. Records, whether bibliographic references, employee records or parts descriptions, have many data elements in varied formats. Because of this, ordering the file (i.e., the group of records) to facilitate retrieval is extremely expensive, if not impossible, even on the most powerful and sophisticated equipment. However, since each index contains only one kind of information it may be ordered relatively easily and in this way tailored to fit the type of data stored for that particular data element. For instance, dates may be indexed in chronological order or in reverse chronological order. Indexing does have economic limits. If many data elements are indexed, the total storage required for indexes may double or triple the amount required for the file itself. This is because of the relatively complex structure of the indexes. Disk storage is also more expensive than tape storage because the mechanism is much more complicated and costly to manufacture.

C. RETRIEVAL

Two examples of manual information retrieval are given as a contrast to computerized information retrieval. In the first example, it is desired to obtain from a personnel file a list of all employees who speak French, have a degree in electrical engineering, have at least two years of professional experience and are not married. The usual practice would be to submit a request for this information to a personnel clerk. This clerk would pull each employee record out of the filing cabinet, one at a time, and examine it to determine if that employee met the conditions of the request. For a large file, this would consume a large amount of the clerk's time in a purely routine task. If the file system is well designed, there might be a list of engineering employees which could be used to reduce the effort. If the personnel department is busy, the requester might have to wait several days to get his information. In addition, one or more employees who meet his requirements might be missed due to human error.

A second example illustrates a retrieval process which is often more wasteful and prone to inaccuracy than the one in the first example. Assume that a medical research scientist wishes to propose the initiation of a new project investigating the effects on human metabolism of the prolonged use of artificial sweeteners. He does not wish to duplicate work which is complete or in progress so he requires information on recent projects in this area. There are several resources he can use in attempting to get this information.

First he can scan all of the applicable journals published during the years he is interested in. Secondly, he may consult his associates to determine if they know of any relevant research. Thirdly, he can contact the leading research organizations to inquire about their current and recent projects. Also, there may be a review published which covers a significant portion of the field. Several major difficulties are inherent in this procedure. It could take several weeks to complete the survey. Several hours effort of highly skilled people is involved. The probability is high that some significant research will be overlooked. A significant amount of the research budget might be consumed in carrying out a function which does not contribute directly to research results.

These difficulties can be alleviated by the use of computerized information storage and retrieval systems. However, it is not necessary, and perhaps not desirable, to have all retrieval functions performed by computer. The user of the system can often benefit, both in terms of the effectiveness and of the economy of retrieval, by having some operations performed manually or by non-computer equipment in conjunction with the computer system. Consider, for example, a bibliographic file, including abstract material or even full text on microfilm. Indexes for the file can be maintained on a computer. The user can then carry out his search through the computer, receiving as output a list of numbers referencing the microfilm which is stored either in cabinets or on special equipment designed for that medium. He might then use a microfilm reader to scan the abstracts and select a final subset of documents. Finally, he or a library assistant would make hard copies of the documents.

The way in which a computer is used to retrieve information from a file depends on several considerations. The first is the frequency with which people request information. Are there several inquiries per day or several per minute? Another consideration concerns the amount of material to be retrieved. Is it normally a yes or no answer (do we have any widgets in stock?), a single name or quantity, a short list of employees and their review dates or a large amount of information such as an address list. A third point is response time: are answers usually required in minutes, hours or days?

The complexity of an inquiry is an involved question and affects, for instance, the way the query is expressed. A SIMPLE REQUEST might be expressed in a single employee name or parts number. A more complex query might be stated in a form which indicates several conditions are to be satisfied before an entry in the file is retrieved. For example, the request "FIND ALL EMPLOYEES WITH SALARY GREATER THAN 10,000 AND AGE LESS THAN 30 AND WITH CLASSIFICATION PROGRAMMER" will return the records of all employees who are programmers under the age of 30 earning more the 10,000 dollars and no other records. This format for a request is called a logical expression.

Another consideration is the complexity of the output. A very simple output consists of every data element in a record, listed in the order it is stored in the file, with one data element per line. A slight complication is introduced if the user specifies that some subset of the elements be listed in a particular order. A sophisticated output facility allows the user to specify page format, i.e., margin size, columnization, double spacing, etc. Some users of the system may require that output be sorted on one or more data elements. For instance, a retrieval request might be for all employees who have an imminent review date with the output listed in order of department number. Often, it is desirable to obtain statistical information on a file which introduces another kind of complexity to the output. For example, what is the average relocation expense claimed by employees hired during the past year or what is the maximum and average number of citations retrieved from the physical science section of a bibliographic file during the last two months.

There are two quite different ways in which a user can communicate with a computer in retrieving information from a file. The first, called BATCH processing, is used when:

1. single requests are for large amounts of information,

2. a response time measured in hours or days is acceptable,
3. output requirements are very complex.

The normal manner of operation for BATCH RETRIEVAL is as follows:

1. a query is formulated and punched on cards,
2. the cards are submitted to a computer operator,
3. he schedules the query and places the cards in a batch with other request cards,
4. the search is executed at the scheduled time (often overnight) and output listed on a high-speed printer,
5. the listing is delivered to the requester.

A purely batch retrieval system is relatively easy and inexpensive to implement but has some definite limitations.

However, an ON-LINE system should be used if the users of the system require answers in minutes or need help from the system in formulating their request, i.e., the first try does not retrieve the material desired and one or more re-formulations must be attempted. In an on-line system several users are communicating with the computer simultaneously. This is accomplished by having many terminals connected to the computer in much the same way that many telephones are connected to a switchboard. In this mode of operation, a retriever enters his request through his terminal and receives a response almost instantaneously. If the request requires a long search, the initial response may be only an indication that the request has been accepted and the computer is in the process of executing it. It may take as long as several minutes to return an answer to some requests. The time that elapses between entering a request and receiving a reply is usually called response time. The elapsed time between receiving a response and entering the next request is normally called think time. People read, reason, and type slowly, in comparison to machine operation time. Think time tends to be fairly long relative to execution time. Thus, the on-line system is able to execute requests for several other users while a single user is digesting the answer to his request.

Basically, there are two types of computer terminals. One type is simply a modified electric typewriter with a wide carriage, a few special function keys and a connection (often a regular telephone line) to the computer. The other type is a screen, similar to the visual part of a television set with a small keyboard added. This kind of terminal is usually called a CRT (short for cathode ray tube) and the output from the computer shown on the screen is called a

DISPLAY. The advantages of a typewriter terminal are: it is relatively inexpensive and it provides hard copy. The disadvantages are: it is relatively slow, it is noisy (especially if several are clustered in one location) and it requires more effort from the user. The advantages of a CRT are: it is virtually noiseless, it is relatively fast (some models can display hundreds of characters in the blink of an eye), and it can be used in ways that make man-machine communication very efficient and effective. The disadvantages are: it provides no hard copy and is expensive. It is possible to combine typewriter and CRT into one terminal and gain a great deal of flexibility but the cost is greater than either device alone.

In many cases, it is not desirable to have a purely batch or a purely on-line information system. Fortunately, there are several ways to combine the two concepts into a single system. The simplest solution is to have an on-line system going during the day and a batch system during the night shift. A more sophisticated solution and one which allows more efficient use of the computer and gives more flexible service to the user community is a system which handles both on-line and batch requests simultaneously. The on-line part of the system has priority and all requests from terminals are satisfied as they are entered. However, the computer frequently runs out of requests to execute and waits for a message to be entered from some terminal. During this wait time, the batch part of the system is given control of the computer and processes part of the batch workload. When a terminal request is entered, control reverts to the on-line part of the system. The batch system is operating in what is called BACKGROUND processing.

As indicated above, both a query and the resulting output can range from very simple to very complex. In order to clarify a discussion of various kinds of retrieval, a brief outline of a session at a terminal follows. The first step that the user takes is to sign on, or "Log On", to the system. This consists of turning on the device and waiting for a signal that the computer is ready for communication. In some cases it is necessary to dial the computer's 'phone number'. The user then keys in a few pieces of general information like his name and account number. The next step is usually the selection of one of the available files. The system then responds with a PROMPT (questions from the computer are called prompts) indicating that it is ready for the user to enter a query.

The user then formulates his query, and types it in. When he hits some particular key (on a typewriter, this is probably the carriage return) the computer examines the message. If it detects an error or does not 'understand' the request, an error message is returned along with a prompt for him to re-enter the query. If the request is correctly formulated, it is placed in a queue (waiting line) and serviced in turn. The queries (and other requests such as output format) are expressed in a language which contains a very limited set of English words and uses a very simple grammatical structure. Since the prompts are considered part of this language and the communication is two way, this language is a CONVERSATIONAL or INTERACTIVE language. Requests directed to a batch system, on the other hand, do not normally have this property.

When the system completes the requested search, it types or displays some response. In the case of certain simple kinds of queries, this message is the requested information. In other cases, the system informs the user of the number of items which meet his CRITERIA (the conditions stated in his query) and waits for him to enter his next request. The user then decides if he wishes to see the information in the retrieved records or if he wishes to refine the criteria and enter a request that will be combined with the previous one to enlarge or reduce the set of retrieved records. An additional step may then be taken; some users will ask for a listing on a high speed printer if he has many pages and wishes to keep a permanent record of his retrieval. The printer is able to list several hundred lines per minute with each line having as many as 133 characters. Also, the printer operates in the background mode and is much less expensive.

The relative simplicity or complexity of retrieval requests, in terms of search and output, determines:

1. the choice of terminal,
2. the way in which files are indexed,
3. the facilities provided for search and output in both the on-line and the batch parts of the system.

For the simplest variety of request, the query contains only the identification of one data element and a single value for it and the output is simply the value of another data element for any record meeting the single criterion. An example of such a request is: RETRIEVE EMPLOYEE JOHN Q. SMITH; OUTPUT SALARY. The system would search the index for employee name, locate the record for John Q. Smith and type or display his salary. For this type of request, there is little difference between a typewriter terminal and a CRT except the cost of the equipment. The complexity increases very little if several items are combined into a LOGICAL EXPRESSION in the search request and more than one item is requested in the output, as: RETRIEVE JOHN Q. SMITH AND HARRY P. ANDERSON; OUTPUT SALARY, POSITION, AGE. There are two distinguishing characteristics of this form of retrieval. The user is able to supply information to retrieve an explicit subset of records from which he requires information. The information he wishes to see is contained in a small number of records in an easily extracted form and he wishes it to be presented essentially as it exists. The principle requirement in this kind of retrieval is that all the data elements which can be specified in a search request must be indexed.

For a contrasting example, consider the query,

FIND ALL TITLES SPIRIT, GHOSTS OR APPARITION,

applied to a file of bibliographic references. The system searches the index for the title data element, locates all references containing any of the three given words in the title and responds with a message indicating how many references have been found, say 46. He then enters the request: OUTPUT TITLE. Suppose the first three titles to be presented were:

The Problem of Ghosts on Television Screens
The Spirit of Christmas
Apparition and Mysticism in Religion.

To reduce the amount of unwanted references in the set he has retrieved, the user enters a modification to his search request: BUT NOT TITLE TELEVISION OR CHRISTMAS OR RELIGION. This might reduce the set to include only relevant material or he might have to make further modifications to the search request. In addition to the problem of retrieving unwanted information, there is also a possibility of not finding some relevant material. There are two things which can be done to alleviate these problems.

Much of the problem of unwanted or lost information is caused by the variety and ambiguity of words in the English language. A contributing factor is that the titles of most books and documents do not reflect completely and accurately the contents. Therefore, searching on the basis of title alone is not an adequate retrieval technique. If a bibliographic file is constructed with a data element that contains phrases descriptive of the subject matter in a document, this data element, when indexed, will usually be useful in retrieval. This type of index is usually called a TOPIC, SUBJECT or KEYWORD index. In addition, an information retrieval system should provide a thesaurus capability. By using a thesaurus a user is able to determine the phrases which are used to describe a topic. He also receives help in formulating his request in a way which helps ensure the retrieval of all relevant material. For instance, if he consults the thesaurus under the word ghost, he might receive the response: SEE ALSO POLTERGEIST.

A third type of retrieval usually has a fairly simple and explicit request in terms of the search but a complex or lengthy requirement for output. For example, in accessing a parts inventory file, to find all parts which are out of stock: RETRIEVE ALL PARTS,

> STOCK = 0; LIST NAME, PART NUMBER,
> ORDER DATE, AVERAGE MONTHLY SALES,
> PRICE; ORDER ALPHABET (NAME).

This request might be entered either through a terminal or, on punched cards, into the batch system. Because of the requirement to sort the output, it would be executed by the batch system. In this example, if there was an index on the data element STOCK, an entry in that index would contain a list of the locations in the file of the records of all parts which were out of stock. Each of these records would be retrieved, the data elements specified for output extracted and an intermediate file created, probably on disk.

This intermediate file would be used as input to a sort program which would produce the output on a high speed printer, ordered alphabetically by part name. If no index existed for the data element STOCK, the batch retrieval would have to read every record in the file and check for a zero value for STOCK.

When a file is set up, a choice is made of the data elements which are to be indexed. Since an index requires a significant amount of storage and adds processing time to the file maintenance, an evaluation is made of the frequency with which that data element might be used as an access point. This helps determine if the cost of the index is justified by expected savings in the processing of queries.

A second example of a retrieval request with output requirements that demand extra processing is the query to a personnel file:

> FIND ALL EMPLOYEES, POSITION SECRETARY; OUTPUT
> AVERAGE AGE, SALARY RANGE, AVERAGE SALARY.

For this request, the system locates the records for all secretaries, computes the average age and salary and lists them along with the lowest and highest secretarial salary. This request could be processed by either the on-line or batch system since the computation is a fairly simple operation.

D. FILE MANAGEMENT

An information storage and retrieval system can support a number of files. For each of these files, there must be someone who is responsible for its management. The person who assumes this responsibility is sometimes called a FILE MANAGER. His tasks include:

1. estimating the size of the file,
2. deciding whether it is to be a direct-access on-line file or a sequential file,
3. specifying the data elements and the indexing requirements,
4. determining who is authorized to access the information contained in it,
5. providing the data for the initial file buildup,
6. supervision of the people who maintain the file.

FILE MAINTENANCE is the process of:

1. adding, deleting and modifying records in the file,
2. editing data to ensure the reliability of the information,
3. initiating the use of backup facilities,
4. executing recovery procedures when damage occurs to the file.

A BACKUP facility provides the ability to make copies of the file on magnetic tape and to maintain a log of recent changes or additions to the file. Together, these may be used to restore a file when some information has been lost or damaged due to computer, program or human malfunction.

The first task of the file manager is FILE DEFINITION, which is the process of specifying the FILE CHARACTERISTICS. Great care should be taken in defining these characteristics since many of the choices made at this time can seriously limit the information which can be put into the file. These choices may restrict and hamper file maintenance tasks. The file manager should take advantage of any consulting services which are offered by the SYSTEM MANAGER, who is responsible for the design, development and maintenance of the information system itself. He may also be in charge of the operation of the computer and related equipment. In fact, in some organizations, his title might be operations manager.

The items which must be specified in the file definition are: the data elements, the properties of the data elements, indexing requirements, thesaurus facilities, display and report formats, editing requirements, partitioning criteria, backup needs and security requirements. Each data element is given a name which is used in the remainder of the definition specifications, in retrieval requests and in output requests. Many systems also allow abbreviations and synonyms for data element names. Other properties to be specified for data elements are DATA TYPE, maximum length and multiplicity. Data type describes the kind of information contained in an element, e.g., numbers, dates, names of people, codes or text. The MAXIMUM LENGTH is the largest number of characters which any value of an element may have and it is used in checking the input data for errors. MULTIPLICITY is simply an indication of whether or not the data element may have more than one value for any given record in the file. Examples of singular data elements are employee name and publisher's address; examples of multiple data elements are languages spoken by an employee and authors of a book.

After considering the various needs of the people who will be retrieving information from the file, the manager must specify the indexing requirements for the file. The first consideration is: which data elements are to be used in expressing search requests? Each of these elements must then be indexed. In addition to indicating the elements to be indexed, he must select which editing facility will be applied to the values in that index. Consider, for example, the title index of a bibliographic file. There are several editing functions which the manager may wish to have performed on titles as they are indexed. First he may wish to delete special characters, such as commas, quotes, periods and colons. Secondly, he may specify a DICTIONARY of words like "IT", "THE", and "A" which should not be indexed. This dictionary is often called an exclusion list; if prepared carefully, it can save considerable storage and processing costs.

For bibliographic files, the manager must specify the contents of a THESAURUS for that file since the words and their relationships are dependent on the subject matter of the file. The thesaurus entry for a word (or a phrase) may have a list of synonyms for that word which helps the user in retrieving further relevant material. It may also show hierarchial relations with other words, i.e., words which are more specific or more general in nature but concerned with the same topic.

While the system will provide some standard formats for display of information on terminals and for listings to be produced on high speed printers, some file managers may wish to specify special formats tailored to the needs associated with their own files. The specification of editing requirements, partitioning criteria, backup needs and security requirements will be described in the appropriate paragraphs below.

The second major task of the file manager is to acquire the data which constitute the information in the file. This data may exist in any of several forms, e.g., file cards, printed material, punched cards or magnetic tape. It may, as in the first two cases above, have to be converted to a form which can be read by the computer. If the data is on cards or magnetic tape, a computer program may have to be written which alters the format so that the input programs of the information system can handle it. Finally, the file manager will have to initiate, with the assistance of the system or operations manager, the process of file building. This normally consists of punching a few system control cards and delivering the input data to a dispatch clerk or a computer operator.

Maintenance of the file includes the functions of adding new information (bibliographic references for recently acquired books), deleting or purging obsolescent material (the records of terminated employees) and the modification of information, (correction of spelling, salary raises, change of address, updating of inventory). For reliability of the file, it is necessary to edit the information as it is input and to provide for backup and recovery. Some editing may be done by the system but much of it can often be done only by manual means. For instance, the computer can be programmed to recognize that JAN 41, 936 is not a legal date but not that the "e" was left off of the name Johnstone. Unfortunately, there are occasions when a computer malfunction or a programming error will cause some information in one or more files to be altered or destroyed. In order to prevent this from becoming a disaster, an information system must provide facilities for backup and recovery. The most common technique used for this purpose consists of periodically copying the file onto a magnetic tape and storing it out of harm's way. In addition, a TRANSACTION FILE is maintained (probably on tape also) of all changes to the file (additions, deletions, etc.) since the last backup was executed. Thus, when damage occurs to an on-line file, recovery is achieved by restoring it from the last backup tape and re-executing the recent changes.

One more very important responsibility of the file manager is prescribing the availability of the file. It may not be economically feasible to have the file on-line all the time the system is operational. So, he may decide to make it available for retrieval only during certain scheduled hours. At other times the disk(s) containing the file can be stored away from the computer. This will free part of the computer equipment for use with other files. Since the access mechanism itself is much more expensive than the disk, a significant savings can be achieved this way. A second availability factor concerns who is able to retrieve from the file. Some files may be public in that any one who has a terminal and an authorized account number may access them. Others may be private with only the file manager and his associates permitted to retrieve information from them. To support this restricted accessibility and to prevent unauthorized persons from altering information in a file, the system must provide a security facility. This usually involves the specification of PASSWORDS by the manager. A user must then know a password to access a private file or to alter the contents of any file.